МАТЕМАТИЧЕСКОЕ ПРОСВЕЩЕНИЕ

Третья серия

выпуск 24

Москва
Издательство МЦНМО
2019

УДК 51.009
ББК 22.1
М34

Адрес редакции:

119002, Москва, Б. Власьевский пер., д. 11, МЦНМО
(с пометкой «Математическое просвещение»)
Email: matpros@yandex.ru
Web page: www.mccme.ru/free-books/matpros.html

М34 **Математическое просвещение.** Третья серия, вып. 24. —
М.: МЦНМО, 2019. — 205 с.
ISBN 978-5-4439-1388-9

В сборниках серии «Математическое просвещение» публикуются материалы о проблемах современной математики, изложенные на доступном для широкой аудитории уровне, статьи по истории математики, обсуждаются проблемы математического образования.

УДК 51.009
ББК 22.1

12+

ISBN 978-5-4439-1388-9

Содержание

Юлию Сергеевичу Ильяшенко — 75!

И. В. Щуров

4 ноября 2018 года исполнилось 75 лет Юлию Сергеевичу Ильяшенко. По протоколу тут должен быть список должностей и регалий юбиляра, и я его начну приводить: доктор физико-математических наук, профессор мехмата МГУ и Корнельского университета (до 2017 года), один из основателей, ректор и профессор Независимого Московского университета;

ординарный профессор Высшей школы экономики; сотрудник Математического института им. В. А. Стеклова, один из главных редакторов Moscow Mathematical Journal… Продолжать можно долго — активность Юлия Сергеевича меня всегда восхищала — но, пожалуй, можно остановиться.

О научных достижения и интересах Юлия Сергеевича тоже можно писать долго. Они включают в себя шестнадцатую проблему Гильберта (Юлий Сергеевич доказал конечность числа предельных циклов полиномиального векторного поля на плоскости), теорию аттракторов (ему принадлежит одно из определений аттрактора — статистический аттрактор), бифуркации динамических систем (совсем недавно Юлий Сергеевич открыл новую, совершенно неожиданную главу этой теории — глобальные бифуркации векторных полей на двумерной сфере), комплексные слоения и многие другие вопросы теории дифференциальных уравнений и смежных областей. Даже перечисление основных работ заняло бы много страниц — но мне вместо этого хотелось бы рассказать несколько личных историй.

В начале 2003 года я был студентом второго курса мехмата МГУ и, имея аж три сессии за плечами, на лекции ходил… иногда. К тому моменту я уже твёрдо усвоил, что прямо на лекции могу понять процентов десять, а остальное можно разве что кое-как записать в надежде разобраться когда-нибудь потом, ближе к сессии (и, скорее всего, по чужим конспектам). Юлий Сергеевич тогда по осенним семестрам работал в Корнеллском университете (США), а весной читал дифференциальные уравнения и спецкурсы на мехмате. Свою часть диффуров (они до этого уже шли семестр) Юлий Сергеевич начал с понятия фазового потока.

«Представьте себе, что вы смотрите, как течёт река. Вы покрасили каждую точку поверхности воды в свой цвет и подождали минуту: каждая точка попала на новое место; отображение, которое на них таким образом подействовало, будет отображением фазового потока за одну минуту». Вместо привычной скороговорки «определение-теорема-доказательство-формула-формула-доказательство-завершено» тут были мотивировки, примеры, иллюстрации, объяснение механизмов. Конечно, определения, теоремы и доказательства тоже были — чёткие, аккуратные и строгие, — но при этом они магическим образом оказывались понятными на лекции. К экзамену я готовился по собственным конспектам — пожалуй, впервые за время учёбы. И понял, что это именно та математика, которая мне нравится.

Не раздумывая ни минуты, я попросился к Юлию Сергеевичу в ученики, а Юлий Сергеевич меня взял. Так я попал в удивительный семинар по динамическим системам, который смело могу назвать своей научной

семьёй. Вернее, сперва я попал на Летнюю школу, о которой следует сказать особо.

Каждое лето Юлий Сергеевич проводит Летнюю школу «Динамические системы», ориентированную на участников семинара и тех, кто интересуется его тематикой. Школа обычно продолжается десять дней и по насыщенности не уступает целому семестру. Обзорные курсы и лекции по отдельным темам, «ликбез» для начинающих, обсуждение новых задач, подготовка статей — на ней происходит много всякой работы. Отдельного упоминания заслуживает гуманитарная составляющая, отражающая широкий кругозор Юлия Сергеевича, — вечерние гуманитарные лекции на самые разные темы, поэтические и музыкальные вечера, «шарады» и песни под гитару. Всё это создаёт удивительно тёплую атмосферу — на мой взгляд, не менее важную для работы, чем научные занятия.

На одной из школ Юлий Сергеевич читал гуманитарную лекцию про историю московской математической школы — начиная от Лузина. Он рассказывал, наверное, полтора часа, и в конце лекции, как обычно, спросил: «Есть ли какие-то вопросы?». Спустя пару мгновений несколько слушателей одновременно спросили: «А что было дальше?». И Юлий Сергеевич продолжил рассказывать…

Летняя школа проходит обычно в Ратмино (под Дубной), но несколько раз мы выезжали в довольно экзотические места: в альплагерь Безенги на Кавказе, на Соловецкие острова и в Словакию. В этих поездках работа школы начинается прямо по пути — если не в виде лекций (их всё-таки сложно вести, например, в поезде), то в виде бесед с учениками и учеников между собой.

По пути на Соловки, на теплоходе, я впервые услышал рассказ Юлия Сергеевича о «чёрном двадцатилетии» мехмата, который произвёл на меня очень большое впечатление[1)]. Вместе с высочайшими стандартами профессиональной деятельности Юлий Сергеевич передаёт своим ученикам свой открытый и гуманистический взгляд на мир, моральные и этические стандарты.

Впрочем, вернёмся к математике. С 2012 года Юлий Сергеевич читает математический анализ на матфаке Высшей школы экономики. Это, казалось бы, классический курс, в котором много десятков лет ничего не менялось — что здесь можно сделать нового? Юлий Сергеевич взялся за него со свойственным ему энтузиазмом и стремлением к совершенству — и обнаружил, что сделать можно многое. Сейчас время от времени

[1)] Подробнее см. в интервью Юлия Сергеевича для polit.ru: https://polit.ru/article/2009/07/28/ilyashenko2/

он делится с нами — его повзрослевшими учениками — своими педагогическими находками. Часть из них направлена на то, чтобы сделать курс доступным для понимания прямо на лекциях. Другая часть — на включение новых элементов и рассказы о связях с разными областями математики — дифференциальными уравнениями, функциональным анализом, теорией особенностей. Усилия не остаются незамеченными — год за годом Юлий Сергеевич получает статус лучшего преподавателя по итогам студенческого голосования.

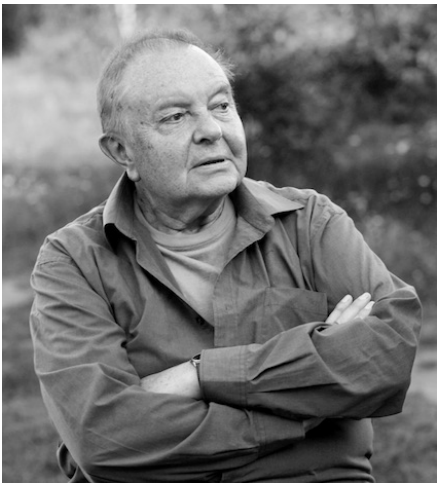
Пожалуй, главное, чему я научился у Юлия Сергеевича, — что лучший учитель — это тот, кто искренне любит своих учеников. Это всегда взаимно.

С днём рожденья, дорогой Юлий Сергеевич! Долгих лет жизни и новых замечательных открытий!

Илья Валерьевич Щуров, НИУ ВШЭ
ilya@schurov.com

Владимир Андреевич Успенский (27.11.1930–27.06.2018)

В. М. Тихомиров



Очерк жизни. Свыше семидесяти лет жизнь Владимира Андреевича Успенского была связана с Московским Университетом. Весной 1945 года он обратился с вопросом к студенту пятого курса мехмата Жене Дынкину. Тот пригласил Володю осенью принять участие в его школьном кружке, и Успенский стал участником этого кружка. Весной следующего года он стал участником IX Московской математической олимпиады. На этой олимпиаде Успенский получил первый приз по восьмым классам.

Эти обстоятельства предопределили выбор высшего учебного заведения. В 1947 году Володя Успенский поступил на механико-математический факультет Московского университета. И в школе, и в университете он учился очень хорошо. В 1949 году произошло важнейшее событие в жизни В. А. Успенского. Вот как (и не раз) он формулировал это: «Меня принял

Редакция благодарит Н. Н. Андреева, предоставившего фотографию В. А. Успенского.

в число своих учеников один из трёх (наряду с Ломоносовым и Менделеевым) великих учёных России Андрей Николаевич Колмогоров».

У Владимира Андреевича Успенского, начиная с его юношеских лет, были очень близкие друзья: со школьных лет — Михаил Константинович Поливанов, с ранних университетских лет — Роланд Львович Добрушин, Вячеслав Всеволодович Ива́нов и Никита Дмитриевна Введенская. В студенческие годы вокруг него сложилась компания ярких и интересных людей — математиков и гуманитариев. Случалось, что Никита Введенская приглашала друзей к себе, иногда некоторых из них, а иногда и всех вместе. Во время дружеских встреч одну из центральных ролей играл Успенский.

Успенский под руководством Колмогорова написал замечательную дипломную работу «Общее определение алгоритмической вычислимости и алгоритмической сводимости». На тему своей дипломной работы Успенский сделал доклад на Московском математическом обществе. Об этом появилось сообщение в журнале «Успехи математических наук».

Затем была аспирантура, завершившаяся защитой кандидатской диссертации «Вычислимые операции над перечислимыми множествам» (1955), признанной Учёным советом выдающейся. Успенский был оставлен в МГУ, а после того, как была образована кафедра математической логики, он стал её сотрудником. С момента поступления в университет началось восхождение на основные для преподавателя ступени: ассистент, доцент, защита докторской диссертации, посвящённой теории вычислимых функций (1963), профессор, заведующий кафедрой. Но эти должности и степени лишь в малой мере отражают тот вклад, который внёс Успенский в науку, просвещение и в жизнь нашего Университета.

Общий обзор творчества. Владимир Андреевич Успенский был таким человеком, интересы которого распространялись на весь интеллектуальный мир. Он был математиком, писал статьи по философии, кибернетике, языковедению, филологии, обладал огромным историческим кругозором. Свойство человека глубоко интересоваться различными сторонами человеческой культуры Колмогоров называл *интеллигентностью*. По этому свойству он ставил Успенского выше всех своих учеников. По широте охвата научных и гуманитарных знаний Успенского трудно с кем-либо сравнить.

Математика. В своих математических исследованиях В. А. Успенский искал философскую сущность изучаемого предмета или явления. Он обдумывал такое фундаментальное понятие, как алгоритм, и исследовал связи понятия алгоритма с частично-рекурсивными функциями. Этому были посвящены начальные работы дипломника и аспиранта Успенского,

приведшие среди разного к понятию алгоритма Колмогорова — Успенского, под этим именем вошедшему в современную литературе по логике.

Всю жизнь Успенского занимали проблемы, навеянные теоремами Гёделя: доказуемость и недоказуемость, вычислимость и невычислимость, а также и более философские вопросы, касающиеся постижимости и непостижимости. Этой тематике Владимир Андреевич посвятил несколько статей, брошюр, книг, включал её в свои основные и специальные курсы.

Успенский читал основной курс по математической логике и спецкурсы «Введение в математическую логику», «Вычислимые функции», «Теорема Гёделя о неполноте», «Язык математики», «Аксиоматический метод».

Двадцать пять учеников Успенского защитили кандидатские диссертации, четверо стали докторами наук.

Колмогоров. Владимир Андреевич посвятил Колмогорову — описанию его личности, творчества и деятельности — значительную долю своих опубликованных работ. Его статья «Колмогоров, каким я его помню» — одна из лучших статей, посвящённых Колмогорову.

Андрей Николаевич был *сеятель*. Он рассеивал фундаментальные идеи за обеденным столом, во время прогулок, на семинарах, на лекциях… Весьма часто собеседник Колмогорова использовал одну из таких идей, и она становилась основополагающей для целого научного направления. Но нередко имя Колмогорова, высказавшего основополагающую идею, нигде не упоминалось. Владимир Андреевич Успенский, наоборот, показывает на своём примере, как некоторые его математические достижения рождались из осмысления замечаний Колмогорова, кратких текстов, написанных на обрывке бумаги, или ответов на вопросы, казалось бы «не на тему». Всё это замечательно описано у Владимира Андреевича в упомянутой статье.

Владимир Андреевич Успенский очень много сделал для пропаганды последнего цикла работ Колмогорова, в котором Андрей Николаевич сделал попытку соединить все линии своей жизни — математическую логику, теорию вероятностей и теорию информации, связав при этом две крайних точки естествознания и философии — хаос и порядок. Иногда основную идею формулируют тремя словами: Колмогоров дал «алгоритмическое определение случайности». Андрей Николаевич оформил этот свой цикл в виде двух, по сути дела последних своих печатных работ 1965 и 1969 годов объёмом около полутора десятков страниц.

Итогом деятельности Успенского по осмыслению этих работ Колмогорова был совместный доклад Колмогорова и Успенского, прочитанный Успенским в сентябре 1986 года в Ташкенте на Всемирном конгрессе общества Бернулли. Как пишет Успенский: «К сожалению, Андрей Нико-

лаевич уже не мог не только ознакомиться с текстом, но даже обсуждать доклад в процессе его подготовки. Разумеется, он полностью основан на его идеях». В. А. Успенскому мы во многом обязаны тем, что эти идеи стали достоянием широкого круга математиков.

Общественная и гуманитарная деятельность. Владимир Андреевич имел очень активный общественный темперамент. Он был подвижником, ставившим перед собой необычайно высокие просвещенческие цели, многие из которых были не только задуманы и объявлены, но и осуществлены. Среди людей, не «встроенных в государственную систему» и так много добившихся на ниве просвещения, Успенского не с кем сопоставить.

Языкознание. Владимир Андреевич является одним из основоположников научного направления в языкознании, получившего название *математической лингвистики*. Это направление у нас имеет точную дату своего рождения. Оно родилось 24 сентября 1956 года. Тогда состоялся первый семинар на филологическом факультете под названием «Некоторые применения математических методов в языкознании» — первый семинар по математической лингвистике в нашей стране. Руководителями семинара были профессор П. С. Кузнецов и два ассистента — В. В. Ива́нов и В. А. Успенский. Двум ассистентам и принадлежала идея организовать этот семинар.

И здесь мимоходом Колмогоров поставил задачу для обсуждения на семинаре, через которую стала проясняться сама основная цель математической лингвистики.

Перед семинаром Успенский зашёл к Колмогорову и рассказал, что идёт на первое заседание семинара по применению математики к языкознанию. Дальнейшее Владимир Андреевич описывает так: «К замыслу семинара Колмогоров отнёсся сочувственно. Он посоветовал предложить участникам семинара две задачи для самостоятельного решения (обе на определение понятия: дать строгие определения понятий ‘ямб’ и ‘падеж’). <…> Что касается падежа, то какое бы то ни было определение этого понятия, хотя бы и неверное, просто отсутствовало».

Для гуманитария сама мысль о том, что на протяжении сотен лет существования грамматик многих языков не было достигнуто чёткого понимания такого основополагающего для большинства языков понятия, как «падеж», была совершенно парадоксальной и революционной. Первое осмысление задачи Колмогорова дал Успенский в статье «К определению падежа по Колмогорову» (1957). А при решении задачи Колмогорова стала постепенно проясняться сама основная цель математической линг-

вистики, которую спустя несколько лет Колмогоров сформулировал так: надо подвергнуть язык «исчерпывающему формальному исследованию современными в смысле логических приёмов методами».

Первый фундаментальный труд, который соответствовал поставленной цели, был выполнен Андреем Анатольевичем Зализняком, представившим в 1965 году диссертацию «Классификация и синтез именных парадигм современного русского языка» на соискание степени кандидата филологических наук. Никому, кроме Успенского, не пришла мысль оценить беспримерную не только в отечественном, но и мировом языкознании работу, присуждением Зализняку за эту работу докторской степени. Приложив усилия, которые доступны только ему, Владимир Андреевич добился поставленной цели, и А. А. Зализняку была присуждена докторская степень. Кстати, решение задачи Колмогорова об определении падежа было окончательно получено Зализняком.

Развитие математической лингвистики постоянно подпитывалось тем, что ныне зовётся информатикой, куда входит, в частности, машинный перевод. Успенский активно участвовал в этой лингво-информационной деятельности.

…В начале шестидесятых годов Роланд Львович Добрушин занимался теорией информации и принял с информационной точки зрения участие в развитии математической лингвистики. Однажды состоялась дружеская пикировка между Добрушиным и Успенским: кого из них следует считать основоположником, а кого классиком этого развивающегося научного направления. Друзья к согласию не пришли, но несомненно, что основоположником является Успенский, а классиками они оба.

Филология. Осознание значимости математической лингвистики и место её в изменившемся мире, в котором определяющую роль стала играть технология с её новейшими средствами приёма и передачи информации и фантастическим развитием компьютеров, побудило Владимира Андреевича Успенского искать новые организационные формы университетского образования лингвистов. Приложив огромные, только для него возможные усилия, Успенский добился организации в МГУ специального отделения, в котором происходило бы объединение лингвистики с математикой. Это отделение называлось сначала Отделением теоретической и прикладной лингвистики, а потом оно было переименовано в Отделение структурной и прикладной лингвистики (сокращённо ОСИПЛ). Успенский разработал программу по математике для этого отделения и организовал преподавание этого предмета на ОСИПЛе. Для всего этого ему пришлось преодолеть воистину непреодолимые препятствия.

Философия. Было сказано, что В. А. Успенский и саму математику рассматривал с философской точки зрения. Ещё в пятидесятые годы он стал печататься в философских изданиях (среди которых «Новая философская энциклопедия», журнал «Вопросы философии», сборник «Закономерности развития современной математики» и другие). В качестве примера приведём статьи Успенского «Семь размышлений на темы философии математики», «Что такое парадокс?», «Абстракция актуальной бесконечности». В. А. Успенский был официальным оппонентом диссертации З. Н. Микеладзе на соискание учёной степени доктора философских наук.

Успенский по просьбе своего учителя участвовал в подготовке статьи Колмогорова «Кибернетика» в 51 томе Большой советской энциклопедии. Эта подготовка нашла своё отражение в работе Ива́нова, Поливанова и Успенского «Тезисы о кибернетике с комментариями». Необычайно интересна переписка Колмогорова с В. А. Успенским и его друзьями — В. В. Ива́новым и М. К. Поливановым. В переписке обсуждается очень широкий спектр философских вопросов.

Лектор и популяризатор. Владимир Андреевич был замечательным лектором, возможно, лучшим среди «математико-гуманитарных» лекторов. В его лекциях и докладах всегда сочеталась суровая математическая определённость с высокими философскими обобщениями. Выступления Успенского всегда были полны юмора, иронии и истинного остроумия.

На протяжении всей своей жизни Владимир Андреевич стремился преодолеть барьер между математическим и гуманитарным. Заметная доля творчества Успенского посвящена популяризации математики. В качестве примера приведу его брошюры: «Некоторые приложения механики к математике», «Треугольник Паскаля», «Машина Поста», «Теорема Гёделя о неполноте», «Простейшие примеры математических доказательств».

В каком-то смысле Владимир Андреевич Успенский завершил дело своей жизни, посвящённое общекультурным и гуманитарным проблемам. Сейчас публикуются работы В. А. Успенского об этих проблемах в пяти книгах, под общим названием «Труды по нематематике». Заглавие не представляется вполне удачным. Кажется более естественным без излишних объяснений написать просто: В. А. Успенский. Мемуары и труды по философии, языкознанию и филологии. Четыре тома уже вышли. Они в свободном доступе по адресу `https://www.mccme.ru/memoria/vau/`. Четыре тома были полностью подготовлены автором к печати, а последний будет издан в этом (2019) году. Пятитомник Успенского подводит итоги его все-

жизненных размышлений о движении человеческой мысли, об истории и о судьбе и творчестве отдельных личностей.

К этому надо присоединить его математические труды и популярные сочинения, не включённые в пятитомник. Из всего этого складывается творчество замечательного учёного Владимира Андреевича Успенского. Читая Успенского, нигде вы не найдёте следов уныния. Всюду вы имеете возможность наблюдать мысли и поступки деятельного и активного человека, который стремится выполнить завет Поэта: «Сотри случайные черты». Он не обещает нам прекрасного мира, но побуждает нас сохранять Веру, Надежду и Любовь во всё хорошее (и ко всему хорошему), что есть в окружающем нас несовершенном мире.

Владимир Михайлович Тихомиров, мехмат МГУ
vmtikh@googlemail.com

В. А. Успенский как историк математики, науки и цивилизации

К статье Александра Шеня «Gauss multiplication trick?»

А. Л. Семёнов

Предлагаемая вниманию читателя статья Александра Шеня посвящена Владимиру Андреевичу Успенскому (ВАУ) не только потому, что автор — ученик Владимира Андреевича. Она отражает важную линию деятельности Успенского в науке и в жизни. Эта линия в большой степени отражена в вышедшем посмертно, но подготовленном Владимиром Андреевичем при жизни пятом, 1100-страничном томе сочинений[1)], как и в других томах его собрания сочинений. (О самом А. Шене можно прочитать на с. 325–331 указанной пятой книги.)

С чисто математической точки зрения содержание публикации Шеня относится к важной вехе в том, что сегодня называется Mathematical Computer Science, а по-русски можно назвать математической информатикой; в рассматриваемый период Колмогоров называл эту область математической кибернетикой. Эта веха — решение А. Карацубой поставленной А. Н. Колмогоровым задачи о вычислительной сложности умножения целых чисел. О самом решении и некоторых его обобщениях и усилениях можно прочитать в статье А. Белова и В. Тихомирова «Сложность алгоритмов»[2)].

Указанное решение Карацубы во многих отечественных и зарубежных публикациях считается одним из первых результатов в теории сложности вычислений, при этом имеющим бесспорное практическое значение. Замечательно, что Колмогоров, последним научным достижением которого

[1)] *Успенский В. А.* Труды по нематематике. 2-е изд., испр. и доп.: В 5 кн. Книга пятая. Воспоминания и наблюдения. М.: Объединённое гуманитарное издательство. Фонд «Математические этюды», 2018.

[2)] *Белов А.*, *Тихомиров В.* Сложность алгоритмов // Квант. 1999. № 2. С. 8–11, http://kvant.mccme.ru/pdf/1999/02/kv0299belov.pdf

стало открытие сложности объектов, заложил основы и второго, ещё более важного направления теории сложности — сложности вычислений, о котором идёт речь.

Понимание Колмогоровым важности задачи видно из того, что он сам записал её решения, найденные молодыми математиками — А. А. Карацубой и Ю. П. Офманом (подробнее об этом см. с. 20).

Своей публикацией Шень тем самым отдаёт дань учителю В. А. Успенского — А. Н. Колмогорову, при том что научными потомками В. А. Успенского являются сам А. Шень и — в некоторых важных аспектах — автор данных строк.

Но есть и ещё одна связь предлагаемой заметки Шеня с деятельностью Успенского. И автору заметки, и автору настоящих строк пришлось принимать участие в исследованиях В. А. Успенского, состоящих в поиске истоков того или иного математического достижения.

Отмечу, что эти поиски часто в качестве своего результата приводили не только к добросовестному цитированию, но и к выяснению оттенков исходной математической мысли, оказывавшихся существенными для математического содержания нашей собственной работы. Успенский был скрупулёзен в выписывании цитат, оформлении ссылок, сравнении переводов с первоисточниками, сопоставлении различных изданий и т. д. При этом часто выявлялись ошибки и несоответствия там, где их никто не ожидал.

В готовящейся в «Успехах математических наук» публикации, посвящённой В. А. Успенскому, обращено внимание на многие случаи, когда открытия Успенского оказывались незамеченными другими, часто из-за того, что и сам Владимир Андреевич их не публиковал должным образом.

Склонность Владимира Андреевича к поиску истоков, объяснению и выявлению в таком поиске чьих-то ошибок, курьёзов и парадоксов, было важным элементом его деятельности как бытописателя и историка. И здесь он не ограничивался математикой. Приведём один из сотен примеров, которые можно найти в сочинениях ВАУ, при этом (с учётом специфики нашего сборника) именно математических, хотя количественно нематематические превалируют (см. цитированную книгу Успенского, с. 813):

> «В качестве вступления в дискуссию о Лобачевском можно также спросить, в чём состоит аксиома о параллельных. Большинство <…> сформулирует эту аксиому так: „через точку, не лежащую на прямой, можно провести прямую, параллельную этой прямой“. На самом деле сформулированное утверждение является не аксиомой, а несложно доказываемой теоремой. Аксиома же о парал-

> лельных состоит в том, что через точку, не лежащую на прямой, можно провести не более одной прямой, параллельной исходной прямой. Причину такого искажения объясняет элементарный филологический анализ. Дело в том, что в средней школе, для простоты, обычно внушают такую формулировку: …можно провести одну и только одну прямую…, не заостряя внимания на том, что оборот „можно провести одну“ выражает здесь теорему, а „можно провести только одну“ — аксиому. В результате в сознании остаётся более простая идея о возможности, а более сложная идея о единственности теряется. Но сказанное никак не объясняет всеобщего заблуждения о сущности сделанного Лобачевским открытия; причины этого заблуждения так и остаются загадкой».

Предлагаемая публикация А. Шеня является продолжением указанной линии деятельности В. А. Успенского и достойной данью памяти нашего учителя. Считаю очень удачным её появление именно в «Математическом просвещении», во многом ориентированном на начинающих математиков. Это может содействовать воспитанию у них серьёзного и уважительного отношения к своим предшественникам, желанию прочитать первоисточники и тщательной подготовке своих работ. Важно и то, что они при этом погрузятся в живую математическую жизнь во всё более актуальной области.

Увлекательного чтения, коллеги!

Алексей Львович Семёнов, академик, зав. кафедрой математической логики и теории алгоритмов МГУ

alsemno@ya.ru

Gauss multiplication trick?

А. Шень

Владимиру Андреевичу Успенскому

§ 1. Недоразумение

В 2014 году в издательстве МЦНМО вышел русский перевод книги [6]; переводил её Александр Куликов (ПОМИ РАН, Петербург), а я был редактором перевода. В этой книге в качестве одного из первых примеров быстрых алгоритмов излагается алгоритм быстрого умножения многозначных чисел. Алгоритм этот сводит умножение $2n$-значных чисел к умножению n-значных. Пусть нам надо умножить два числа из $2n$ битов (мы рассматриваем двоичные числа, но это не принципиально). Разобьём двоичные записи $2n$-битовых сомножителей x и y на две половины по n знаков:

$$x = 2^n x_1 + x_0, \quad y = 2^n y_1 + y_0$$

(сдвиг на n двоичных разрядов соответствует умножению на 2^n; все четыре половинки x_1, x_0, y_1, y_0 содержат по n битов каждая). Тогда

$$xy = 2^{2n} x_1 y_1 + 2^n (x_0 y_1 + y_0 x_1) + x_0 y_0.$$

Мы свели нашу задачу к умножению четырёх пар n-битовых чисел $x_1 y_1$, $x_0 y_1$, $y_0 x_1$, $x_0 y_0$ — к четырём задачам половинного размера (умножение на степени двойки, т. е. дописывание нулей, а также операции сложения мы не считаем, так как они проще — сравните сложение и умножение столбиком). Повторяя это сведéние ещё раз, мы получим на следующем шаге 16 задач вчетверо меньшего размера и так далее, пока не придём к однозначным числам, где задача тривиальна.

Несложно понять, что этот подход сам по себе не даёт выигрыша. Если всё подсчитать аккуратно, то общее число операций при умножении n-значных чисел будет расти пропорционально n^2 — ровно так же, как при обычном алгоритме умножения столбиком, где мы умножаем

каждый из n разрядов первого числа на каждый из n разрядов второго. (Увеличение размера вдвое соответствует увеличению числа операций при умножении столбиком вчетверо.)

Но мы получим более быстрый алгоритм, если заметим, что можно обойтись *тремя* умножениями n-битовых чисел вместо четырёх. А именно, если за одно умножение вычислить $(x_1+x_0)(y_1+y_0)$, то потом можно вычесть x_1y_1 и x_0y_0, которые всё равно нужно вычислять, и получить сразу сумму $x_0y_1 + y_0x_1$, не вычисляя каждое из произведений по отдельности. Делая так на каждом шаге рекурсии, мы достигаем существенной экономии: вместо n^2 получается

$$n^{\log_2 3} \approx n^{1{,}59},$$

что заметно даже для не очень больших значений n. Этот более быстрый алгоритм является стандартным алгоритмом умножения в библиотеках работы с многозначными числами[1)].

Его[2)] придумал Анатолий Алексеевич Карацуба (1937–2008) осенью 1960 года; алгоритм был опубликован в 1962 году в [22]. Подробности этой истории рассказал сам автор в [20]:

> Осенью 1960 г. в Московском университете на механико-математическом факультете начал работать семинар по математическим вопросам кибернетики под руководством А. Н. Колмогорова, где А. Н. Колмогоровым была сформулирована гипотеза n^2 ⟨про порядок роста числа битовых операций, необходимых при умножении n-значных чисел[3)]⟩ и поставлен ряд задач об оценке сложности решений линейных систем уравнений и других сходных вычислений. Я активно стал размышлять над гипотезой n^2 и ровно через неделю обнаружил, что алгоритм, которым я надеялся получить нижнюю ⟨так в статье⟩ оценку величины $M(n)$, даёт оценку вида
>
> $$M(n) = O(n^{\log_2 3}), \quad \log_2 3 = 1{,}5849\ldots$$

[1)] Впоследствии были придуманы и другие алгоритмы, более быстрые (асимптотически, т. е. для достаточно длинных чисел); о некоторых из них можно прочитать в [19].

[2)] На самом деле в статье [22] рассматривается задача о возведении в квадрат многозначных чисел — частный случай умножения, к которому легко сводится общий с помощью формулы

$$ab = \frac{(a+b)^2 - a^2 - b^2}{2}.$$

Для этого частного случая сведение к трём задачам меньшего размера выглядит немного иначе: чтобы вычислить $(2^n a + b)^2 = 2^{2n}a^2 + 2^n \cdot 2ab + b^2$, достаточно возвести a и b в квадрат и потом найти $2ab = (a+b)^2 - a^2 - b^2$, сделав ещё одно возведение в квадрат.

[3)] Здесь и далее угловыми скобками отмечены комментарии, не являющиеся частью цитаты.

> После очередного заседания семинара я сообщил А. Н. Колмогорову о новом алгоритме умножения и об опровержении гипотезы n^2. Это сильно взволновало А. Н. Колмогорова, так как противоречило его довольно правдоподобной гипотезе. На следующем заседании семинара мой метод умножения был рассказан самим А. Н. Колмогоровым, и на этом семинар прекратил свою работу. Позднее, в 1962 г., А. Н. Колмогоров написал (может быть, при участии Ю. П. Офмана[4]) небольшую статью и опубликовал её в Докладах АН СССР. Статья называлась так: А. Карацуба, Ю. Офман, Умножение многозначных чисел на автоматах (ДАН СССР, 1962, т. 145, № 2, с. 293–294). Об этой статье я узнал только тогда, когда мне были даны её оттиски. Необычность способа публикации подчёркивается и тем, что обе статьи [5] и [8] ⟨соответственно [24] и [22] в нашем списке⟩ представлены А. Н. Колмогоровым к опубликованию одновременно 13.II.1962.

Удивительным образом (как заметил Куликов) имя Карацубы вовсе не упоминалось в переводимой книге, а идея о возможности сокращения числа умножений с четырёх до трёх приписывалась Гауссу [6, с. 45 (55 в электронном варианте)]:

> The mathematician Carl Friedrich Gauss (1777–1855) once noticed that although the product of two complex numbers
>
> $$(a + bi)(c + di) = ac - bd + (bc + ad)i$$
>
> seems to involve *four* real-number multiplications, it can in fact be done with just *three*: ac, bd, and $(a + b)(c + d)$, since
>
> $$bc + ad = (a + b)(c + d) - ac - bd.$$
>
> ⟨…⟩ this modest improvement becomes very significant *when applied recursively*.

(Здесь речь идёт об умножении комплексных чисел, но разница лишь в знаке перед bd.) Мы с Куликовым подготовили соответствующее примечание: после слов «Время работы соответствующего алгоритма» должна была следовать сноска «Его предложил А. А. Карацуба (ДАН СССР, 1962, т. 145, с. 293–294). — *Прим. перев.*». К сожалению, в процессе подготовки оригинал-макета это добавление было забыто (и теперь должно дожидаться переиздания[5], когда и если таковое будет), на что обратила внима-

[4] Юрий Петрович Офман (родился в 1939 году), ученик Колмогорова, занимался теорией сложности вычислений и теорией автоматов. Автор книги [25].

[5] Книга переиздана в 2019 г., ошибка исправлена.

ние (в своём письме в издательство, пересланном нам с Куликовым 6 августа 2016 года) дочь А. А. Карацубы, Екатерина Анатольевна Карацуба:

> В изданном вами переводе книги «Алгоритмы» С. Дасгупта, Х. Пападимитриу, У. Вазирани содержится фактическая ошибка. Первый в мире быстрый алгоритм, алгоритм быстрого умножения, открытый А. А. Карацубой (задача была поставлена А. Н. Колмогоровым), в этой книге приписан Гауссу, безо всяких ссылок на какие-либо документальные источники и свидетельства.
>
> Американские авторы и раньше пытались приписать идею метода Карацубы (названную Шёнхаге «дивайд энд конкур» ⟨divide and conquer, англ.⟩) кому-то другому, так же как и создание первого быстрого алгоритма. Однако у них нет и не было никаких подтверждающих их страстное желание приписать этот метод какому-то более для них предпочитаемому автору документов (лишь словесные мифы). Нет ни одного быстрого алгоритма в каком-либо тексте, изданном раньше 1963 г. (А. А. Карацуба создал свой алгоритм и рассказал его на колмогоровском семинаре в 1960 г., издан в журнале «Доклады АН СССР» и переведён на английский в 1962 г., рассказан Колмогоровым на многих международных конференциях, начиная с 1960 г., включая международный математический конгресс в Стокгольме в 1962 г., опубликован в США отдельно по стокгольмской лекции Колмогорова в книге-сборнике лекций конгресса в 1963 г.).
>
> Издавать в России (безо всяких комментариев) книгу, в которой воруется основное российское открытие в области вычислительной математики, алгоритм, внедрённый в виде софт- и хардвэр в основные компьютеры мира — это как-то не очень порядочно. А что вы об этом думаете?
>
> С уважением, Е. А. Карацуба

Мне как редактору перевода оставалось только принести извинения за допущенную оплошность[6)] — но стало интересно: что вообще пишут разные авторы об истории вопроса, были ли у Карацубы предшествен-

[6)] Вот ответное письмо:

Добрый день, Екатерина Анатольевна!

Редактор издательства, Юрий Николаевич Торхов, переслал мне Ваше письмо, и я отвечаю Вам в качестве редактора перевода книги Дасгупта, Пападимитриу и Вазирани «Алгоритмы». Да, разумеется, алгоритм умножения по половинам был предложен А. А. Карацубой, и мы с переводчиком даже подготовили соответствующее примечание: после слов «Время работы соответствующего алгоритма» должна была следовать сноска «Его предложил А. А. Карацуба (ДАН СССР, 1962, т. 145, с. 293–294). — *Прим. перев.*». К сожале-

ники, и обнаружилось много любопытных вещей, иллюстрирующих, помимо прочего, пути распространения информации и ошибок в ней. К сожалению, это «расследование» осталось незаконченным — может быть, кто-то из тех, кто увидит этот текст, сможет довести его до конца.

§ 2. Почему Гаусс?

Переводимая книга не давала никаких ссылок на источники, и я попытался выяснить, что пишут на эту тему другие авторы. Поиск в интернете показывает, что Гаусс как изобретатель способа умножения комплексных чисел с помощью трёх умножений действительных чисел упоминается разными авторами. Например, Мур и Мертенс в своей известной книге [12, с. 37] (2011) пишут:

> The first $O(n^{\log_2 3})$ algorithm for multiplying n-digit integers was found in 1962 by Karatsuba and Ofman [447]. However, the fact that we can reduce the number of multiplications from four to three goes back to Gauss! He noticed that in order to calculate the product of two complex numbers, $(a+bi)(c+di) = (ac-bd)+(ad+bc)i$ we only need three real multiplications, such as ac, bd, and $(a+c)(b+d)$, since we can get the real and imaginary parts by adding and substracting these products. The idea of [447] is then to replace i with $10^{n/2}$, and to apply this trick recursively.

Здесь «[447]» — это ссылка на оригинальную статью [22]; в ней были сформулированы два результата с указанием авторства (Ю. П. Офман

нию, по моей вине в процессе подготовки оригинал-макета это добавление было забыто. Если будет переиздание, эту ошибку мы исправим.

Всего хорошего,

Александр Шень

P. S. Прошу прощения за задержку с ответом: мне стало интересно, какова история вопроса и откуда взялось название «Gauss trick». Не будучи специалистом по истории науки, я тем не менее предпринял некоторые любительские разыскания, но к успеху они не привели: действительно, название Gauss trick во многих местах встречается в применении к умножению комплексных чисел за три вещественных умножения, но авторы не приводят соответствующих ссылок, а приведённая в Википедии ссылка на Кнута (Knuth D., The Art of Computer programming, vol. 2, 1998, pp. 519, 706, см. `https://en.wikipedia.org/wiki/Multiplication_algorithm#Karatsuba_multiplication` по состоянию на 9 июня 2016) вводит в заблуждение: Гаусс вообще в этом томе в связи с быстрым умножением не упоминается, а в связи с обсуждаемым алгоритмом умножения даётся ссылка на работу А. А. Карацубы (естественно). Само умножение комплексных чисел за три вещественных умножения обсуждается, но без упоминания Гаусса. (Конец письма.)

для одного из них, А. А. Карацуба для второго, того са́мого алгоритма быстрого умножения многозначных чисел). Но ссылок на Гаусса снова не приводится. Нет их и в другой публикации, упоминающей Гаусса как предшественника Карацубы [16]:

> Indeed, when it comes to multiplying two numbers, the best (or fastest) way to do it is often far from obvious.
>
> One particularly intriguing and efficient multiplication algorithm was developed in the late 1950s by Anatolii Alexeevich Karatsuba, now at the Steklov Institute of Mathematics in Moscow.
>
> Karatsuba's "divide-and-conquer" multiplication algorithm has its roots in a method that Carl Friedrich Gauss (1777–1855) introduced involving the multiplication of complex numbers.
>
> ⟨объясняется, как сэкономить одно умножение⟩
>
> So, Gauss optimization saves one multiplication out of four.
>
> Karatsuba's divide-and-conquer multiplication algorithm takes advantage of this saving. ⟨…⟩
>
> Karatsuba's insight was to apply Gauss optimization to this divide-conquer-and-glue approach, replacing some multiplications with extra additions. For large numbers, decimal or binary, Karatsuba's algorithm is remarkably efficient.

Ссылка на Гаусса (наряду с корректной ссылкой на алгоритм Карацубы и его первую публикацию) была в английской Википедии, когда я туда посмотрел (январь 2017 — в текущей версии этого нет, поскольку я внёс соответствующие исправления в текст), но с ней совсем странная история. Там говорилось [13]:

> **Gauss's complex multiplication algorithm**
>
> Complex multiplication normally involves four multiplications and two additions.
>
> $$(a + bi)(c + di) = (ac - bd) + (bc + ad)i.$$
>
> ⟨…⟩ By 1805 Gauss had discovered a way of reducing the number of multiplications to three [11].

Ссылка «[11]» гласила: "Knuth, Donald E. (1988), The Art of Computer Programming, volume 2: Seminumerical algorithms, Addison-Wesley, pp. 519, 706", и, судя по описанию, имелась в виду классическая книга Кнута [10]. Но указанный год (1988) не соответствует ни второму изданию (около 1981), ни третьему (около 1998). Об алгоритме Карацубы в книге Кнута говорится следующее (Section 4.3.3, с. 294 третьего издания, во втором издании этот текст есть на с. 278):

***4.3.3. How Fast Can We Multiply?**

⟨…⟩ let us consider the following question: *Does every general computer algorithm for multiplying two n-place numbers require an execution time proportional to n^2, as n increases?*

⟨…⟩ The answer to the question above is, rather surprisingly, "No," and, in fact, it is not very difficult to see why. ⟨…⟩ If we have two $2n$-bit numbers $u = (u_{2n-1}\dots u_1u_0)_2$ and $v = (v_{2n-1}\dots v_1v_0)_2$, we can write

$$u = 2^nU_1 + U_0, \quad v = 2^nV_1 + V_0 \tag{1}$$

where $U_1 = (u_{2n-1}\dots u_n)_2$ is the "most significant half" of the number u and $U_0 = (u_{n-1}\dots u_0)_2$ is the "least significant half"; similarly $V_1 = (v_{2n-1}\dots v_n)_2$ and $V_0 = (v_{n-1}\dots v_0)_2$. Now we have

$$uv = (2^{2n} + 2^n)U_1V_1 + 2^n(U_1 - U_0)(V_0 - V_1) + (2^n + 1)U_0V_0. \tag{2}$$

This formula reduces the problem of multiplying $2n$-bit numbers to three multiplications of n-bit numbers, namely, U_1V_1, $(U_1-U_0)(V_1-V_0)$, and U_0V_0, plus some simple shifting and adding operations.

⟨…⟩ the main advantage of (2) is that we can use it to define a recursive process for multiplication that is significantly faster than the familiar order-n^2 method when n is large: If $T(n)$ is the time required to perform multiplication of n-bit numbers, we have

$$T(2n) \leqslant 3T(n) + cn \tag{3}$$

for some constant c. ⟨…⟩ the running time for multiplication can be reduced from order n^2 to order $n^{\lg 3} \approx n^{1.585}$, so the recursive method is much faster than the traditional method when n is large.

⟨…⟩ (A similar but slightly more complicated method for doing multiplication with running time of order $n^{\lg 3}$ was apparently first suggested by A. Karatsuba in Doklady Akad. Nauk SSSR, **145** (1962), 293–294. ⟨…⟩ Curiously, this idea does not seem to have been discovered before 1962; none of the "calculating prodigies" who have become famous for their ability to multiply large numbers mentally have been reported to use any such method, although formula (2) adapted to decimal notation would seem to lead to a reasonably easy way to multiply eight-digit numbers in one's head.)

Ни о Гауссе, ни об умножении комплексных чисел здесь не говорится[7]. Но на указанных в Википедии страницах (519, 706) умножение комплекс-

[7] Предметный указатель книги Кнута перечисляет упоминания Гаусса на страницах 20, 101, 363, 417, 422, 449, 578, 679, 685, 688, 701. Ни одна из этих страниц не содержит никаких упоминаний приписываемого ему способа умножения комплексных чисел.

ных чисел действительно упоминается. На с. 519 (во втором издании с. 501) имеется упражнение (к разделу 4.6.4):

> **41.** [*22*] Show that real and imaginary parts of $(a+bi)(c+di)$ can be obtained by doing 3 multiplications and 5 additions of real numbers, where two of the additions involve a and b only.

А на с. 706 (раздел «Answer to exercises»; во втором издании на с. 647) даётся ответ к этому упражнению:

> **41.** $a(c+d)-(a+b)d+i(a(c+d)+(b-a)c)$. $\langle\ldots\rangle$ Without the restriction on additions there are other possibilities. For example, the symmetric formula $ac-bd+i((a+b)(c+d)-ac-bd)$ was suggested by Peter Ungar in 1963. $\langle\ldots\rangle$ See I. Munro, *STOC* **3** (1971), 40–44; S. Winograd, *Linear Algebra and its Applications*, **4** (1971), 381–388.

Ссылок на Унгара у Кнута тоже нет (и в любом случае алгоритм Карацубы опубликован раньше 1963 года)[8]. В статьях Винограда и Мунро, ссылки на которые приводит Кнут, доказывается, что меньше трёх умножений не получится. В статье Винограда [18] про это сказано так:

> ABSTRACT
>
> The two main results of this note are:
>
> (i) The minimum number of multiplications required to multiply two 2×2 matrices is seven.
>
> (ii) The minimum number of multiplications/divisions required to multiply two complex numbers is three.
>
> $\langle\ldots\rangle$ we note that it is possible to compute a complex product using only three multiplications. For example,
>
> $$ac-bd=ac-bd,$$
> $$ad+bc=(a+b)(c+d)-ac-bd.$$
>
> So the three products which are formed are ac, bd, $(a+b)(c+d)$.

Ссылок на источник формулы у него нет, как и у Мунро, который пишет [14]:

> The Multiplication of Complex Numbers
>
> If the number of multiplications required for a computation is regarded as a measure of its difficulty and these computations are performed using complex numbers, it is natural to ask how many real multiplications are necessary to evaluate the real and imaginary parts

[8] Анатолий Воробей в июле 2018 года написал Унгару с просьбой прояснить ситуацию, и он в ответном письме объяснил, что действительно предложил такой способ в 1960-х годах (“I did have the idea in the mid-sixties and told a few people at New York University and I think to Vinograd too, but I did not know Knuth credited me with it.”)

of a complex product. The natural way of forming a complex product requires four real multiplications. It may, however, be done in three but not in two multiplications

$$(a+bi)(c+di) = (ac-bd)+(ad+bc)i$$
$$\begin{array}{ccc} a(c+d) & - & d(a+b) = ac-bd \\ (1) & & (2) \\ a(c+d) & + & c(b-a) = ad+bc \end{array}$$

Theorem. The evaluation of the product of two complex numbers requires three real multiplications, even if multiplication by real constants is not counted.

Унгар упоминается и в некоторых других публикациях. Например, в [8] написано:

1. Introduction. How many real multiplications are required to multiply two complex numbers? In view of the familiar identity

$$z = (a+ib)(c+id) = ac-bd+i(ad+bc),$$

the answer may appear to be four. However, it is possible to make do with three multiplications, because

$$z = ac-bd+i[(a+b)(c+d)-ac-bd]. \qquad (1.1)$$

This formula was suggested by Peter Ungar in 1963, according to Knuth [14, p. 647].

Ссылка «[14]» указывает на второе издание книги Кнута (где процитированное выше решение упражнения 41 находится на с. 647), но прямых ссылок на Унгара нет.

Таким образом, появление Гаусса в Википедии остаётся загадочным. Согласно данным с сайта, раздел про умножение комплексных чисел «по Гауссу» появился впервые в версии 22:55, 27 May 2009 (в версии 21:24, 22 May 2009 его ещё не было); соответствующая правка — с той же ссылкой на Кнута, что и сейчас — была внесена участником Dmcq (`https://en.wikipedia.org/wiki/User_talk:Dmcq`; сейчас (2017) по этому адресу указывается “Semi-retired. This user is no longer very active on Wikipedia.”). Ссылка на алгоритм Карацубы там была до этого (с первой же версии статьи, 15 июня 2002).

Похоже, что тем не менее эта запись в Википедии привела к распространению недоразумений. Скажем, в [2] (2014) Гаусс также упоминается со ссылкой на Кнута:

It is well known too, that the complex multiplication can be carried out using only three real multiplications and five real additions, because ⟨…⟩

$$(a+jb)(c+jd) = ac - bd + j[(a+b)(c+d) - ac - bd]. \qquad (4)$$

Expression (4) is well known as Gauss' trick for multiplication of complex numbers [17].

Здесь «[17]» — это второе издание книги Кнута [10].

Это «словесное квипрокво»[9] огорчительно, тем более когда вообще алгоритм Карацубы приписывается Гауссу (как, например, в [17]).

§ 3. Бэббидж, divide et impera

Занимаясь этими разысканиями, я обнаружил (благодаря ссылке в [12]) удивительную цитату из Бэббиджа, разрабатывавшего (в XIX веке!) программируемую вычислительную машину (так и не реализованную «в железе»). Он называл её «Analytical Engine». Говоря об умножении многозначных чисел с её помощью, Бэббидж пишет [1, с. 61; в публикации 1864 года с. 125]:

…Thus if $a \cdot 10^{50} + b$ and $a' \cdot 10^{50} + b'$ are two numbers each of less than a hundred places of figures, then each can be expressed upon two columns of fifty figures, and a, b, a', b' are less than fifty places of figures ⟨…⟩ The product of two such numbers is

$$aa'10^{100} + (ab' + a'b)10^{50} + bb'.$$

This expression contains four pairs of factors, $aa', ab', a'b, bb'$, each factor of which has less than fifty places of figures. Each multiplication can therefore be executed in the Engine. The time, however, of multiplying two numbers, each consisting of any number of digits between fifty and one hundred, will be nearly four times as long as that of two such numbers of less then fifty places of figures ⟨…⟩

Thus it appears that whatever may be the number of digits the Analytical Engine is capable of holding, if it is required to make all the computations with k times that number of digits, then it can be executed by the same Engine, but in an amount of time equal to k^2 the former.

[9] Термин, предложенный в [26, 27] Владимиром Андреевичем Успенским, большим знатоком подобных историй о недоразумениях и ошибках восприятия [28, 29] — он обнаружил, среди прочего, что памятник, воспринимаемый многими парижанами как памятник погибшей принцессе Диане, в действительности не имеет к ней никакого отношения!

Перевод:

> …Таким образом, если два числа $a \cdot 10^{50} + b$ и $a' \cdot 10^{50} + b'$ состоят менее чем из ста цифр каждое, то оба числа могут быть разбиты на две части по пятьдесят цифр ⟨буквально: записаны в двух столбцах из пятидесяти цифр⟩; числа a, b, a', b' содержат до пятидесяти разрядов… Произведение двух таких чисел равно
>
> $$aa'10^{100} + (ab' + a'b)10^{50} + bb'.$$
>
> Это выражение содержит четыре пары сомножителей aa', ab', $a'b$, bb', и каждый сомножитель содержит до пятидесяти цифр. Таким образом, Машина сможет выполнить эти умножения. Однако время умножения двух чисел, содержащих от пятидесяти до ста цифр каждое, будет примерно в четыре раза больше, чем время умножения чисел до пятидесяти цифр…
>
> Получается, что каково бы ни было количество цифр в числах, помещающихся в Аналитическую Машину, при необходимости та же Машина может выполнять вычисления и с числами, в которых в k раз больше цифр, но это потребует в k^2 раз большего времени.

Видно, что хотя у Бэббиджа совершенно ясно изложена схема divide et impera (латинское выражение, по-английски говорят «divide and conquer», по-русски обычно переводят это как «разделяй и властвуй» — в данном случае мы делим числа на две части), но он не подозревает о возможности замены четырёх умножений на три и соответственного сокращения времени вычисления — так что даже если предположить, что в каких-то бумагах Гаусса и была подобная идея, то видно, что распространения она не получила.

Отметим ещё классический пример алгоритма типа «разделяй и властвуй», который тоже был придуман в докомпьютерную эру — алгоритм сортировки слиянием (подлежащие сортировке объекты произвольно делятся на две группы; каждая из групп отдельно сортируется, а потом группы сливаются с сохранением порядка). Как пишет Кнут [11, с. 385],

> The idea of merging goes back to another card-walloping machine, the *collator*, which was a much later ⟨в сравнении с машинами, сортирующими карты сначала по одной колонке, потом по другой (radix-sort)⟩ invention (1938). With its two feeding stations, it could merge two sorted decks of cards into one, in only one pass; the technique for doing this was clearly explained in the first IBM collator manual (April 1939). [See James W. Bruce, *U. S. Patent 2189024* (1940).]

Модель «IBM 77 electric punched card collator», разработанная фирмой IBM в 1937 году, описывается так [9]:

> As a filing machine, the Type 77 fed and compared simultaneously two groups of punched cards: records already in file and records to be filed. These two groups were merged in correct numerical or alphabetical sequence. ⟨…⟩ Introduced in 1937, the IBM 77 collator rented for $80 a month. It was capable of handling 240 cards a minute ⟨…⟩ IBM withdrew the Type 77 from marketing on November 27, 1957.

§ 4. Снова о Гауссе

Вопрос о том, откуда взялась версия о Гауссе и трёх умножениях, так и остаётся непонятным. Можно предположить, что всё-таки в каких-то рукописях Гаусса такое замечание имеется (что, разумеется, никак не отменяет бесспорного приоритета Карацубы по части алгоритма быстрого умножения). Но это сейчас кажется мне малоправдоподобным, поскольку никаких подтверждающих упоминаний найти не удалось. Другой вариант, может быть, более правдоподобный — что это результат смешения двух историй: быстрого умножения и быстрого преобразования Фурье.

Преобразование Фурье (в интересующем нас конечном варианте) можно описать как вычисление n значений многочлена $P(x)$ степени меньше n во всех корнях степени n из единицы. Интерполяционная формула Лагранжа говорит, что имеется взаимно однозначное соответствие между наборами коэффициентов и наборами значений, и алгоритм быстрого преобразования Фурье позволяет вычислить это преобразование (в любую сторону) за $O(n \log n)$ действий. Этот алгоритм тоже основан на сведении задачи к меньшей, если n есть степень двойки. А именно, пусть $n = 2k$ и ζ — корень из единицы, порождающий все остальные. Мы хотим вычислить $P(1)$, $P(\zeta)$, $P(\zeta^2)$, …, $P(\zeta^{2k-1})$, где $P(z)$ — многочлен степени меньше $2k$. Если сгруппировать в нём чётные и нечётные члены порознь, то получится $P(z) = P_0(z^2) + zP_1(z^2)$, где P_0 и P_1 — многочлены степени меньше k. Таким образом, нам достаточно вычислить значения многочленов P_0 и P_1 в точках

$$1,\ \zeta^2,\ \zeta^4,\ \ldots,\ \zeta^{2k-2},\ \zeta^{2k} = 1,\ \zeta^{2k+2} = \zeta^2,\ \zeta^{2k+4} = \zeta^4,\ \ldots,\ \zeta^{4k-2} = \zeta^{2k-2},$$

которые являются квадратами корней степени $2k$, т. е. в корнях степени k (каждый встречается дважды). Мы свели задачу к двум задачам вдвое меньшего размера и $O(n)$ умножениям и сложениям (нужным для соединения результатов). Рекурсивное применение этого алгоритма даёт оцен-

ку в $O(n \log n)$ арифметических операций для n, являющихся степенями двойки.

Этот алгоритм был опубликован в статье Кули и Тьюки [5] в 1965 году. Он оказался очень важным с практической точки зрения (непосредственным поводом к их работе была компьютерная обработка сигналов, в частности, данных о волнах в земной коре после ядерных испытаний). Вскоре после публикации обнаружилось, что этот алгоритм неоднократно использовался и публиковался и раньше [4]. Более того, впоследствии выяснилось, что по существу эта же идея содержалась (и использовалась) в записях Гаусса, видимо, относящихся к 1805 году и опубликованных в 1866 году — но написанных на современной Гауссу версии латыни, см. [3, 7], и потому мало кому понятных в настоящее время.

Может быть, эти две истории смешались в чьём-то сознании? Тем более что преобразование Фурье оказалось полезным для быстрого умножения многочленов (вычислим значения в корнях из единицы, перемножим их за $O(n)$ действий, а потом сделаем обратное преобразование), что в свою очередь позволило улучшить оценку Карацубы (алгоритм Шёнхаге — Штрассена, 1971: двоичная запись числа по существу есть значение многочлена с коэффициентами 0 и 1 в точке 2 в конечном поле).

В любом случае, хорошо бы по возможности уменьшить путаницу в этом деле, независимо от причины, по которой она возникла…

§ 5. Дополнение

Алексей Устинов, член редколлегии «Математического просвещения», задал вопрос про «трюк Гаусса» на сайте MathOverflow [23]. Отвечая на этот вопрос, Карло Бенакер предпринял библиографические разыскания, которые также не привели к цели (выяснению того, почему описанный способ умножения комплексных чисел приписывают Гауссу) — зато он обнаружил чуть более ранний текст [15], записки лекций Папаконстантину в университете Йорка 2005 года, где этот способ также приписывается Гауссу (но снова без конкретной ссылки).

Список литературы

[1] *Babbage C.* On the principles and development of the calculator and other seminal writings / Ed. by P. Morrison and E. Morrison. Dover publications, 1961. Более ранняя публикация (1864): *Babbage C.* Passages from the life of a philosopher. Longman et al. London, 1974.
https://books.google.ru/books?id=Fa1JAAAAMAAJ&pg=PA125 (с. 125).

[2] *Cariow A, Cariowa G.* A Hardware-oriented Algorithm for Complex-Valued Constant Matrix-Vector Multiplication. https://arxiv.org/pdf/1410.6937v1.pdf

[3] *Cooley J. W.* The Re-Discovery of the Fast Fourier Transform Algorithm // Mikrochimica Acta [Wien]. 1987. III. P. 33–45. См. также: *Cooley J. W.* How the FFT Gained Acceptance // HSNC'87 Proceedings of the ACM Conference on History of scientific and numeric computation. Princeton, NJ, USA, May 13–15, 1987. ACM Publishers.

[4] *Cooley J. W.*, *Lewis P. A. W.*, *Welsh P. D.* Historical Notes on the Fast Fourier Transform // IEEE Transactions on Audio and Electroacoustics. 1967. Vol. 15, iss. 2. P. 76–79. DOI:10.1109/TAU.1967.1161903

[5] *Cooley J. W.*, *Tukey J. W.* An Algorithm for the Machine Calculation of Complex Fourier Series // Mathematics of Computation. 1965. Vol. 19, № 90. P. 297–301. https://www.jstor.org/stable/2003354

[6] *Dasgupta S.*, *Papadimitriou C. H.*, *Vazirani U. V.* Algorithms. McGraw-Hill, 2008. Copyright notice on the electronic draft: 2006. Глава 2: https://people.eecs.berkeley.edu/~vazirani/algorithms/chap2.pdf (Рус. перев.: *Дасгупта С.*, *Пападимитриу Х.*, *Вазирани У.* Алгоритмы / Пер. с англ. А. Куликова под ред. А. Шеня. М.: МЦНМО, 2014. 320 с.

[7] *Heideman M. T.*, *Johnson D. H.*, *Burrus C. S.* Gauss and the History of the Fast Fourier Transform // IEEE ASSP magazine. 1984. Vol. 1, iss. 4. P. 14–21. https://ieeexplore.ieee.org/document/1162257, http://www.cis.rit.edu/class/simg716/Gauss_History_FFT.pdf

[8] *Higham N. J.* Stability of a method for multiplying complex matrices with three real matrix multiplications // SIAM J. Matrix Anal. Appl. 1992. Vol. 13, № 3. P. 681–687. https://pdfs.semanticscholar.org/fa55/3f9528a38cba2a23986071354425ea748480.pdf

[9] International Business Machines (IBM). *IBM 77 electric punch collator*. http://www-03.ibm.com/ibm/history/exhibits/vintage/vintage_4506VV4004.html

[10] *Knuth D. E.* The Art of Computer Programming. Vol. 2: Seminumerical algorithms. Third edition. Addison-Wesley (copyright: 1998, first printing: September 1997). ISBN 0-201-89684-2. (First edition: 1969; second edition: 1981).

[11] *Knuth D. E.* The Art of Computer Programming. Vol. 3: Sorting and Searching. Second edition. Addison-Wesley (copyright: 1998, first printing: March 1998). ISBN 0-201-89685-0.

[12] *Moore C*, *Mertens S.* The Nature of Computation. Oxford University press, 2011.

[13] Multiplication algorithm // Wikipedia, https://en.wikipedia.org/wiki/Multiplication_algorithm, 04.01.2017.

[14] *Munro I.* Some results concerning efficient and optimal algorithms // Proceedings of the Third Annual ACM Symposium on Theory of Computing (STOC 1971). ACM, 1971. P. 40–44.

[15] *Papakonstantinou P. A.* (York University). Introduction to Divide and Conquer. Integer multiplication faster than $O(n^2)$. https://www.eecs.yorku.ca/course_archive/2005-06/F/3101/dc_intro.pdf

[16] Ivars Peterson. Divide and Conquer Multiplication. *Science News*. February 11, 2007. https://www.sciencenews.org/article/divide-and-conquer-multiplication

[17] Tim Roughgarden. Video lectures, CS161 — Design and Analysis of Algorithms. Lecture 9 of 172. http://openclassroom.stanford.edu/MainFolder/VideoPage.php?course=IntroToAlgorithms&video=CS161L1P9

[18] *Winograd S.* On Multiplication of 2 × 2 Matrices // Linear Algebra and its Applications. 1971. Vol. 4. P. 381–388. http://www.sciencedirect.com/science/article/pii/0024379571900097

[19] *Белов А.*, *Тихомиров В.* Сложность алгоритмов // Квант. 1999. № 2. С. 8–11. http://kvant.mccme.ru/pdf/1999/02/kv0299belov.pdf

[20] *Карацуба А. А.* Сложность вычислений // Труды Математического института РАН. 1995. Т. 211. С. 186–202. http://www.ccas.ru/personal/karatsuba/divcru.pdf

[21] *Карацуба А. А.* Комментарии к моим работам, написанные мной самим. (Подготовили к публикации (частичной) С. А. Гриценко и Е. А. Карацуба) // Современные проблемы математики. 2013. Вып. 17. С. 7–29. DOI:http://dx.doi.org/10.4213/spm41

[22] *Карацуба А. А.*, *Офман Ю. П.* Умножение многозначных чисел на автоматах // ДАН СССР. 1962. Т. 145, № 2. С. 293–294.

[23] Обсуждение вопроса Алексея Устинова на сайте MathOverflow. https://mathoverflow.net/questions/319559/gauss-trick-vs-karatsuba-multiplication/319589

[24] *Офман Ю. П.* Об алгоритмической сложности дискретных функций // ДАН СССР. 1962. Т. 145, № 1. С. 48–51.

[25] *Офман Ю.* О христианстве и иудаизме. М.: Изд-во ПСТГУ, 2015.

[26] *Успенский В. А.* Почему на клетке слона написано «буйвол»: Наблюдения о словесных квипрокво (подменах текста) и их причинах // Труды по **не**математике. Кн. 4: Филология. М.: Объединённое гуманитарное издательство. Фонд «Математические этюды», 2012. С. 254–461.

[27] *Успенский В. А.* Ещё раз о словесных квипрокво // Труды по **не**математике. Кн. 5: Воспоминания и наблюдения. М.: Объединённое гуманитарное издательство. Фонд «Математические этюды», 2018. С. 800–807.

[28] *Успенский В. А.* Парижские сюрпризы // Труды по **не**математике. Кн. 5: Воспоминания и наблюдения. М.: Объединённое гуманитарное издательство. Фонд «Математические этюды», 2018. С. 616–622.

[29] *Успенский В. А.* Привычные вывихи // Труды по **не**математике. Кн. 5: Воспоминания и наблюдения. М.: Объединённое гуманитарное издательство. Фонд «Математические этюды», 2018. С. 808–816.

Александр Шень, LIRMM CNRS & University of Montpellier, ИППИ РАН
sasha.shen@gmail.com

Заметки с Международного конгресса математиков

А. Ю. Окуньков

Летом 2018 года в Бразилии, в Рио-де-Жанейро, прошёл очередной, уже 28-й по счёту, Международный конгресс математиков (сокращённо МКМ, а по-английски ICM). Конгрессы математиков проводятся с 1897 года и являются не только важнейшими форумами в нашей математической профессии, но и старейшими научными мероприятиями подобного формата. Сейчас они проводятся под эгидой Международного математического союза (ММС). На МКМ вручаются медали Филдса и другие награды Международного математического союза, а именно премии Неванлинны, Гаусса и Лилавати, а также медаль Черна.

Костяк программы МКМ составляют пленарные и секционные доклады, которые предоставляют участникам МКМ уникальную возможность увидеть широчайшую панораму последних достижений во всех областях математики. Пленарные и секционные докладчики тщательно подбираются программным комитетом ММС и обычно очень добросовестно готовят свои доклады, вкладывая в них много мысли и души. На многочисленных формальных и неформальных площадках МКМ математики всего мира знакомятся друг с другом, делятся своими проблемами и опытом, учатся друг у друга, вдохновляют друг друга и т. д. Наконец, публичные лекции конгресса несут математическое знание в поистине широкие массы, которые помимо участников конгресса включают любителей математики, школьников и студентов.

Мне посчастливилось быть участником МКМ в Рио-де-Жанейро, и я бы хотел поделиться своими впечатлениями и мыслями с читателями «Математического просвещения» в этих заметках. Сразу оговорюсь, что современная математика настолько широка и глубока, что целиком вместить её в голову не под силу никому. Не могу утверждать даже, что в моей собственной относительно узкой специальности я детально понял все доклады. Поэтому математику от меня далёкую, а таковая составляет

подавляющую часть, я могу пересказывать только очень крупными мазками. Единственная цель, которую подобный пересказ может преследовать, это быть своего рода красочной открыткой из далёкой манящей страны. Я очень надеюсь, что мой пересказ заинтересует читателя и даст ему стимул отправиться в неблизкое путешествие в ту самую манящую страну, а менее иносказательно — засесть за специализированную литературу.

Ключевым и с нетерпением ожидаемым моментом дня открытия конгресса является вручение медалей Филдса, которые считаются самой высокой наградой в математике. В 2018 году медалей Филдса были удостоены Каушер Биркар, Акшай Венкатеш, Алессио Фигалли и Петер Шольце. По традиции церемония включает небольшие лекции о заслугах лауреатов (в английском языке для них используется заимствованное из латыни слово laudatio, что может быть переведено как похвала или восхваление). О заслугах новоиспечённых лауреатов рассказали Кристофер Хэкон, Питер Сарнак, Луис Каффарелли и Михаэль Рапопорт соответственно. Опираясь на эти laudatio и на пленарные доклады самих лауреатов, я попробую рассказать о достижениях новых филдсовских медалистов.

Начать рассказ, видимо, правильно с *Петера Шольце*, ибо даже в такой звёздной компании он стоит особняком. Если имена других медалистов для кого-то вероятно были сюрпризом, то награждение Шольце предсказывали и ожидали все. Трудно не почувствовать восхищения, даже ошеломления, как от глубины, так и от количества открытий, сделанных Шольце, при том что на момент награждения ему было всего 30 лет. При этом работает он в области, которая мне лично представляется невероятно тяжёлой и технической, но вот как-то у него получается находить правильные точки зрения на вопросы, ставившие в тупик старшие поколения математиков.

Математический анализ и дифференциальная геометрия оперируют с вещественными (и комплексными) числами, которые получаются из поля рациональных чисел $\mathbb{Q}$ (или поля $\mathbb{Q}(\sqrt{-1})$ в случае комплексных чисел) с помощью процедуры пополнения относительно обычной нормы $|x|$. В теории чисел исключительно важно рассматривать p-адические нормирования, в которых число x тем меньше, чем на бóльшую степень простого числа p оно делится. Это позволяет подходить к теоретико-числовым вопросам, таким как делимость, аналитически. Над полями, полными относительно p-адического нормирования, можно развивать аналог комплексной геометрии, и это даёт как бы геометрический подход ко многим ключевым проблемам теории чисел. Другое дело, что и в смысле геометрической интуиции, техники, да и объективного существа феноменов, p-адическая ситуация заметно хитрее комплексной.

Одной из первых ярких идей Шольце было понятие перфектоидного поля и перфектоидного пространства. От перфектоидного поля $\Bbbk$ требуется:

1) полнота относительно p-адического нормирования $|\cdot|$,
2) плотность образа отображения $|\cdot|\colon \Bbbk \to \mathbb{R}_{>0}$,
3) существование некоторых корней p-й степени.

Например, если мы к обычному p-адическому полю $\mathbb{Q}_p$ присоединим корни $\sqrt[p^n]{p}$ для всех $n = 1, 2, \dots$ и затем снова p-адически пополним, то получим перфектоидное поле характеристики 0. А если вместо этого мы рассмотрим формальные ряды с членами вида t^{1/p^n} и коэффициентами в конечном поле $\mathbb{F}_p$, то это будет пример перфектоидного поля характеристики p. Отталкиваясь от этого определения, Шольце вводит понятия перфектоидной алгебры и перфектоидного пространства. Несмотря на свою угрожающую нетёровость, эти объекты сразу проявили себя как самое мощное техническое средство в p-адической геометрии. Многие казавшиеся совершенно неприступными гипотезы (выдвинутые такими математиками, как Делинь, Тейт, Хохстер и др.) были в течение нескольких лет или полностью, или в очень большой общности доказаны Шольце и другими (из которых, видимо, следует особенно выделить Баргава Бхатта и Ива Андре (Ives André)) с использованием техники перфектоидных пространств.

Многие из подобных гипотез касаются когомологий p-адических пространств, и не будет преувеличением сказать, что Шольце и его соавторы научили математиков думать о них совсем по-новому и, как мне кажется, гораздо яснее и понятнее. В привычной геометрии группы когомологий являются, вероятно, самым основным алгебраическим объектом, который можно сопоставить комплексному многообразию, и много важной информации течёт по этому мосту между алгеброй и геометрией в обоих направлениях. К примеру, в задачах геометрической теории представлений само пространство представления часто реализуется как некоторая группа когомологий. В том же духе когомологии некоторых специальных p-адических пространств играют ключевую роль в программе Ленглендса, и результаты Шольце, среди прочего, помогают строить по ним представления Галуа в гораздо большей общности, чем было известно ранее.

Когомологии комплексного многообразия можно вычислять очень по-разному: можно сосчитать когомологии Чеха для подходящего открытого покрытия, а можно использовать теорию де Рама и дифференциальные формы (которые могут быть определены алгебраически для алгебраических многообразий). Согласованность этих конструкций очень неочевидна. С одной стороны, интегралы форм по циклам, которые тут участвуют, могут быть очень интересными трансцендентными числами,

как, например, число π. С другой стороны, интегралы форм по циклам игнорируют кручение, т. е. элементы конечного порядка в когомологиях, а кручение часто несёт в себе очень важную информацию.

В алгебраической и теоретико-числовой ситуации вместо старых когомологий Чеха мы имеем этальные когомологии, и внимание большого числа специалистов было сосредоточено на сравнении их с когомологиями де Рама и другими теориями когомологий. Гротендик учил, что в основе их всех должна лежать одна универсальная теория, а именно теория мотивов. Завершение теории мотивов предполагает, однако, доказательство гипотез типа гипотез Ходжа и Тейта, к которым математики вот уже очень долгое время никак не могут подобрать ключей. Тем более замечательно, что Шольце вместе со своими соавторами Бхаттом и Морроу придумал некоторую теорию когомологий, которая зависит от дополнительных параметров и при определённом выборе их значений приводит и к этальным, и к де-рамовским, и к кристальным когомологиям, включая кручение. Ещё более замечательно, что в основе их построения лежит замечательная конкретная деформация комплекса де Рама для аффинного пространства, а именно его q-разностная деформация, столь любимая в теории специальных функций и теории представлений. Просто вместо частных производных мы берём

$$\nabla_{i,q} f(\dots, x_i, \dots) = \frac{f(\dots, qx_i, \dots) - f(\dots, x_i, \dots)}{(q-1)x_i}.$$

При $q \to 1$ это превращается в обычную производную, а лишний параметр q как раз и даёт теории Бхатта, Морроу и Шольце возможность включить многие теории как частные случаи.

Отсчитывающий свою историю с работ Гейне 1840-х годов, q-разностный анализ переживал периоды как высокого, так и среднего интереса со стороны других областей математики. Один подобный всплеск интереса был вызван открытием и исследованием квантовых групп в работах школы Фаддева, Дринфельда и Киотской школы. Другой всплеск приходится на наше время. С одной стороны, как показали Бхатт, Морроу и Шольце, q-разностная техника справляется со многими глубочайшими вопросами в теории чисел. С другой стороны, можно сказать, что квантовые группы и связанная с ними математическая физика шагнули в следующее измерение или даже в следующие измерения. Подробный рассказ об этом отвлёк бы нас слишком далеко, но хочу констатировать тот факт, что роль q-разностных уравнений в математической физике и связанной с ней алгебраической геометрии и теории представлений тоже внезапно выросла.

Каушер Биркар был удостоен медали Филдса за свои работы по многомерной комплексной алгебраической геометрии. Золотым стандартом в алгебраической геометрии является теория гладких алгебраических кривых или компактных римановых поверхностей. Важнейшим инвариантом кривой C является её род $g(C)$, который, в духе уже обсуждавшейся нами эквивалентности между различными построениями когомологий, можно определить или как число ручек у римановой поверхности C, или как число линейно независимых дифференциальных форм на C. Алгебраические, геометрические и теоретико-числовые свойства C исключительно сильно зависят от того, в какой из трёх следующих классов она попадает.

1) Имеется единственная гладкая полная кривая рода 0, это комплексная проективная прямая $C = \mathbb{P}^1$, или сфера Римана. На ней нет ненулевых всюду регулярных сечений расслоения Ω^1_C дифференциальных форм, которое в данном случае совпадает с каноническим линейным расслоением ω_C. Более того, двойственное касательное к C расслоение $\omega_C^{-1} = T_C$ обильно, т. е. имеет много сечений. Поэтому $C = \mathbb{P}^1$ есть простейший пример многообразия Фано. Напомним, что каноническое линейное расслоение для гладкого многообразия X определяется как $\omega_X = \Lambda^{\dim X}\Omega^1_X$, т. е. как старшая внешняя степень кокасательного расслоения X.

2) Для кривых рода $g > 1$, наоборот, некоторая степень ω_C обильна, тем самым они являются простейшим примером многообразий общего типа. Таких кривых много, их классы изоморфизма параметризуются некоторым комплексным многообразием размерности $3g - 3 > 0$, к которому мы ещё вернёмся позже.

3) Промежуточное положение занимают кривые рода 1, для них расслоение ω_C тривиально, т. е. на них имеется единственный с точностью до множителя дифференциал, который нигде не вырожден. В большей размерности тривиальность ω_C означает существование всюду невырожденной формы старшей степени. Такие многообразия принято называть многообразиями Калаби — Яу.

Цель многомерной алгебраической геометрии — развить подобного рода понимание для алгебраических многообразий произвольной размерности. Ясно, что картина будет неизмеримо сложнее по целому ряду причин. Уже начиная с размерности 2 классификация с точностью до изоморфизма является слишком утончённой для большинства задач. Действительно, если у нас есть гладкая алгебраическая поверхность, то её можно раздуть в любой точке и получить новую, неизомофную, но очень близкородственную поверхность. Разумно поэтому расширить отношение эквивалентности до бирациональной эквивалентности, которая означает, что два многообразия имеют изоморфные открытые по Зарискому

подмножества, или, что то же самое, изоморфные поля рациональных функций. Для гладких полных кривых бирациональная эквивалентность равносильна изоморфизму.

Далее, на произвольном алгебраическом многообразии X ни одно из линейных расслоений $\omega_X^{\pm 1}$ не обязано быть обильным, например, X может расслаиваться над базой общего типа со слоями, которые являются многообразиями Фано или Калаби — Яу. Задачей программы минимальных моделей является сведение произвольного X к такого рода структурам посредством некоторой контролируемой последовательности бирациональных преобразований. Это целая отрасль математики, в развитие которой внесли огромный вклад как зарубежные учёные, такие как Мори, Коллар, Хэкон и др., так и отечественная школа алгебраической геометрии, и в особенности В. А. Исковских и В. В. Шокуров.

Среди результатов *Биркара* можно, пожалуй, выделить два. Первый, полученный совместно с Хэконом, Маккернаном и Кассини, говорит о том, что так называемое каноническое кольцо, т. е. кольцо, образованное сечениями всех степеней ω_X, конечно порождено для гладкого проективного многообразия X над полем комплексных чисел. Таким образом, в частности, можно говорить о его проективном спектре X_{can}, что есть так называемая каноническая модель X. Эта теорема есть яркий заключительный аккорд в долгом развитии идей многих замечательных математиков. По моему скромному суждению, Хэкона стоило бы отметить медалью Филдса за это и другие его достижения, пока ему ещё не было сорока лет. Хоть жизнь и сложилась иначе, признание всё же нашло и Хэкона, и Маккернана в виде Премии за прорыв в математике (Breakthrough prize in mathematics), вручённой им в 2017 году.

Второй же, теперь уже сольный результат Биркара касается классификации многообразий Фано. В любой фиксированной размерности > 1 может существовать бесконечно много различных гладких многообразий Фано, например раздутия проективной плоскости $\mathbb{P}^2$ в $\leqslant 8$ точках. Однако все они ограничены в том смысле, что образуют конечное число семейств (параметризуемых положением точек раздутия в предыдущем примере). Ограниченность трёхмерных гладких многообразий Фано была доказана В. А. Исковских в 1970-х годах, а случай произвольной размерности был завершён Наделем, Кампаной, Колларом, Мори и Мияокой в начале 1990-х годов.

Простые примеры показывают, что если отказаться от гладкости, то ограниченность пропадает. Однако если сузить допустимый класс особенностей с помощью некоторой численной характеристики ε, то для любого заданного $\varepsilon > 0$ так называемые ε-лог-терминальные многообразия Фано в любой заданной размерности снова ограничены. Это утверждение было

высказано как гипотеза В. Алексеевым и братьями А. и Л. Борисовыми и стало известно как BAB-гипотеза. В своём laudatio Хэкон охарактеризовал эту гипотезу как самую важную гипотезу о многообразиях Фано. Частные случаи этой гипотезы были известны благодаря работам самих Алексеева и Борисовых, Каваматы, Хэкона, Маккернана и Шу. Биркару же удалось доказать эту гипотезу в самой полной общности. Этот его результат и принёс ему славу.

Как и в случае Петера Шольце, о работах *Акшая Венкатеша* на конгрессе рассказывал его бывший научный руководитель — Питер Сарнак. Его задача, полагаю, была столь же сложна, как и задача Михаэля Рапопорта, рассказывавшего о достижениях Петера Шольце. Спектр научных интересов самого Сарнака уже сам по себе практически необъятен, а его ученик Венкатеш способен генерировать свежие идеи и получать первоклассные результаты в столь разнообразных областях, что просто дух захватывает. Тут и динамические системы, и автоморфные формы, и масса других важнейших объектов современной теории чисел, которые все в его статьях живут в сложно переплетённом симбиозе.

Важная тема, подчёркнутая Сарнаком в его выступлении, — это цикл задач о равнораспределении, связанных с так называемыми субконвексными оценками на автоморфные L-функции. Задачи такого рода имеют длинную историю в теории чисел, но особенный прогресс в их анализе был достигнут замечательным отечественным математиком Ю. В. Линником и его последователями (сам Венкатеш в своих статьях постоянно подчёркивает влияние идей школы Линника). В простейшем примере речь идёт о целых точках на сфере или гиперболоиде размера $\sqrt{d}$, т. е. о множестве S_d решений уравнения

$$x_1^2 + x_2^2 \pm x_3^2 = d, \quad (x_1, x_2, x_3) \in \mathbb{Z}^3,$$

и о распределении соответствующих точек $|d|^{-1/2}(x_1, x_2, x_3)$ на единичной сфере или единичном гиперболоиде. В работах Линника и Скубенко было доказано, при некоторых предположениях на d, что точки $|d|^{-1/2}S_d$ становятся равнораспределены при $d \to \infty$ относительно естественных $G(\mathbb{R})$-инвариантных мер на сфере и гиперболоиде соответственно. Здесь G есть группа матриц, сохраняющих квадратичную форму $x_1^2 + x_2^2 \pm x_3^2$; она и её подгруппы играют очень важную роль в анализе данной задачи. Собственно, Линник и был первым, кто понял важность идей, связанных с действиями групп и эргодической теорией в данном контексте.

В восьмидесятых годах, т. е. через тридцать лет после работ Линника, его результаты были усилены Дюком следующим образом. Предположим

для простоты, что речь идёт о сфере. Асимптотическая равнораспределённость конечных наборов точек S_d означает слабую сходимость соответствующих дискретных вероятностных мер $\frac{1}{|S_d|}\sum_{x\in S_d}\delta_x$ к заданной мере μ, а это, в свою очередь, эквивалентно тому, что

$$\langle f\rangle_d = \frac{1}{|S_d|}\sum_{x\in S_d} f(x) \to 0 \qquad (1)$$

для плотного множества таких тестовых функций, что $\int f d\mu = 0$. Уже Линник использовал результаты Б. А. Венкова, из которых следует связь между S_d и группой классов идеалов в кольце целых поля $\mathbb{Q}(\sqrt{-d})$. Развивая эту связь, можно получить, для правильно подобранных f, точное выражение для $\langle f\rangle_d$ через специальные значения L-функций. Напомним, что L-функции — это очень глубокие по своим свойствам функции комплексной переменной s, представляющие собой далеко идущие обобщения ζ-функции Римана

$$\zeta(s) = \sum_{n>0}\frac{1}{n^s}, \quad \operatorname{Re} s > 1,$$

впервые рассмотренной Эйлером в Санкт-Петербурге в 1740 году. Гораздо более общие L-функции могут быть построены, например, по автоморфному представлению редуктивной группы над некоторым полем. Утверждение (1) тогда является следствием некоторой так называемой субконвексной оценки для L-функций, которая для $\zeta(s)$ превращается в оценку $\zeta(1/2+it) \ll (1+|t|)^{\varepsilon}$ на критической линии $\operatorname{Re} s = 1/2$, где $\varepsilon < 1/4$.

Если бы мы доказали гипотезу Римана (а ещё лучше — обобщённую гипотезу Римана для всех L-функций!), то из неё следовала бы подобная оценка с любым $\varepsilon > 0$, а также много других замечательных следствий, однако это, видимо, дело далёкого будущего. Пока специалисты по автоморфным формам доказывают субконвексные оценки в разных специальных случаях, и это целая большая область современной теории чисел. В частности, Венкатеш внёс в неё важный вклад, завершив, совместно с Мишелем, доказательство общей субконвексной оценки $L(\pi, s)$ для всех автоморфных представлений π группы $\mathrm{GL}(2)$ над числовым полем.

Следующий после сфер и гиперболоидов случай связан с кубическими полями и может быть интерпретирован в терминах асимптотической равнораспределённости замкнутых орбит для действия группы диагональных матриц $H \subset \mathrm{SL}(3,\mathbb{R})$ на пространстве $X_3 = \mathrm{SL}(3,\mathbb{R})/\mathrm{SL}(3,\mathbb{Z})$. На каждой такой замкнутой орбите живёт конечная H-инвариантная мера, и теорема Айнзидлера, Венкатеша, Линденштраусса и Мишеля утверждает, что некоторые пакеты таких орбит асимптотически равноспределены в X_3. Прямая

оценка общих средних типа $\langle f \rangle_d$ потребовала бы тут новых труднодостижимых субконвексных оценок. Вместо этого авторы резко сужают запас необходимых f, используя мощные структурные результаты о всех вообще возможных конечных H-инвариантных мерах, полученные Айнзидлером, Катком и Линденштраусcом. После этого всё равно остаётся много работы, прежде чем эта замечательная мозаика из эргодической теории, гармонического анализа и теории чисел принимает свою окончательную форму.

Интересно заметить, что другое направление исследований Венкатеша, которому был посвящён его собственный доклад на конгрессе, опять связано с кручением в когомологиях, хоть и не совсем с той же стороны, что и работы Шольце. Можно, наверное, сказать, что 2018-й был хорошим годом для кручения в когомологиях. А если серьёзно, то речь тут идёт вот о чём. Пусть G обозначает вещественную полупростую группу Ли, например $\mathrm{SL}(n, \mathbb{R})$, а $K \subset G$ обозначает максимальную компактную подгруппу в ней. В $\mathrm{SL}(n, \mathbb{R})$ это будет $\mathrm{SO}(n, \mathbb{R})$ с точностью до сопряжения. Многообразие G/K обладает замечательной G-инвариантной римановой метрикой, для $n = 2$ в нашем примере получится плоскость Лобачевского. Пусть теперь $\Gamma \subset G$ есть арифметическая подгруппа, например $\Gamma = \mathrm{SL}(n, \mathbb{Z}) \subset \mathrm{SL}(n, \mathbb{R})$. Многообразие (или, в общем случае, орбифолд) двойных классов смежности $X = \Gamma \backslash G / K$ является одним из главных геометрических объектов современной теории чисел. В частности, его когомологии играют важную роль и перерабатываются, согласно ви́дению Ленглендса, вместе с действующими на них соответствиями Гекке в некоторые представления Галуа. Когомологии X особенно хорошо изучены, когда G/K является эрмитовым симметрическим пространством, что в нашем примере отвечает случаю $n = 2$. Если обозначить через δ так называемый дефект, т. е. разницу между рангами G и K, то эрмитов случай отвечает $\delta = 0$.

Если взять убывающую последовательность нормальных подгрупп

$$\Gamma_1 \supset \Gamma_2 \supset \ldots \supset \Gamma_N \supset \ldots, \quad \bigcap \Gamma_i = \{1\},$$

то размерность $H^k(X_N, \mathbb{C})$ растёт пропорционально объёму X_N в случае $\delta = 0$ и средних когомологий, а в остальных случаях медленнее. Это показали де Джордж и Воллах ещё в конце 1970-х годов. Венкатеш и Бержерон установили аналогичный результат для кручения в $H^k(X_N, \mathbb{Z})$. В этом случае естественно делить логарифм порядка подгруппы кручения на объём X_N, и этот предел оказывается ненулевым только при $\delta = 1$ и $k = \frac{\dim}{2} + 1$. Чтобы лучше понять это огромное кручение, Венкатеш вводит дополнительные операторы типа Гекке, которые рождают новые классы кручения из имеющихся, что также помогает прояснить, хотя бы гипотетически, как себя проявляют эти классы кручения на стороне представлений Галуа.

Наконец, четвёртый лауреат, *Алессио Фигалли*, — это уже чистый аналитик, главные работы которого посвящены кругу задач типа задачи Монжа — Канторовича, также известной как задача об оптимальной транспортировке. В задаче Монжа ищется отображение $f\colon X \to Y$, переводящее заданную меру μ_X в заданную меру μ_Y и минимизирующее функционал вида

$$C(f) = \int\limits_X c(x, f(x))\, d\mu_X,$$

обычно называемый функцией стоимости. В хозяйственной интерпретации функция $c(x, f(x))$ может быть стоимостью перевозки единицы некоторого продукта из точки x в точку $f(x)$. Если из каждой точки $x \in X$ разрешить везти не в одну заданную точку $f(x) \in Y$, а распределить имеющийся в x продукт по Y согласно некоторой мере ϕ на $X \times Y$, то получится задача Канторовича. От ϕ требуется иметь заданные проекции μ_X и μ_Y и минимизировать стоимость $\int_{X\times Y} c(x, y)\, d\phi$. Задача Канторовича есть задача линейного программирования, поэтому она проще. Но по смыслу и сути она очень близка к задаче Монжа и очень помогает в анализе последней. Много ярких результатов в общем круге задач Монжа — Канторовича были получены отечественной школой, в том числе А. Д. Александровым, А. М. Вершиком, Н. В. Крыловым, В. А. Рохлиным, А. В. Погореловым, Р. Л. Добрушиным, Ю. В. Прохоровым и многими другими, а среди зарубежных учёных нельзя не упомянуть Л. Амброзио, Я. Бренье, С. Виллани, Э. Калаби, Л. Каффарелли, Р. Маккэна, Ш. Яу, и этот список тоже можно долго продолжать.

При всей, казалось бы, утилитарности транспортной задачи, она относится не к периферии, а к самому центру математики. Одну из ролей оптимальной транспортировки можно, вероятно, сравнить с ролью конформных отображений в анализе на плоскости. Если, например, для доказательства какой-то оценки надо сравнить два множества X и Y (или две меры на них), то очень полезно рассмотреть соответствующее отображение f и исследовать его свойства. Конкретный пример мы увидим позже, а со множеством других примеров можно познакомиться в относительно недавнем обзоре В. И. Богачева и А. В. Колесникова[1)].

Ключевому вопросу о регулярности отображения f были посвящены фундаментальные работы многих перечисленных и неперечисленных математиков. Важной вехой в его решении стали работы Л. Каффарелли, который и выступал с laudatio работ Алессио Фигалли. В своём докладе

[1)] *Богачев В. И., Колесников А. В.* Задача Монжа — Канторовича: достижения, связи и перспективы // УМН. 2012. Т. 67, вып. 5(407). С. 3–110.

Каффарелли особенно подчеркнул важность соболевской регулярности $f \in W^{2,1}$, доказанной Фигалли совместно с де Филипписом. В другой статье той же серии де Филиппис и Фигалли доказывают стабильность, т. е. непрерывную зависимость f от условий задачи в соответствующих пространствах Соболева. Поскольку, как отмечалось выше, рассмотрение f полезно в доказательстве многих неравенств, из этого выводится стабильность соответствующих оценок.

Хотелось бы упомянуть одно приложение оптимальной транспортировки к задачам теории вероятности из совместной работы Фигалли и Алисы Гионе. Классический вопрос теории вероятности состоит в том, как будет себя вести маленький фрагмент большой случайной системы. Например, можно рассмотреть столь малый объём газа или другой большой системы взаимодействующих частиц, что ожидаемое число частиц в нём конечно, и изучать полученную меру на конфигурациях частиц. В качестве взаимодействующих частиц могут выступать, например, собственные числа случайной эрмитовой матрицы (по некоторым историческим причинам этот случай пользуется особой популярностью). Естественное предположение состоит в том, что наблюдаемое в малом объёме зависит не от всех деталей поведения большой системы, а от конечного набора макроскопических параметров типа плотности и температуры. В системах с локальным или достаточно быстро убывающим взаимодействием этот феномен можно описывать в формализме мер Гиббса, развитом Р. Л. Добрушиным и его школой. Хотя взаимодействие собственных чисел случайной матрицы и очень нелокально, тем не менее детали поведения всей большой системы стираются также и для случайных матриц, и для многих аналогичных систем. Люди, занимающиеся случайными матрицами, называют этот феномен универсальностью, и много работ посвящено его доказательству при различных предположениях. Идея, я смею предположить, выдвинутая Гионе, состояла в том, что можно случайный набор частиц перевести в модельный набор с помощью некоторой оптимальной транспортировки f, а затем уже исследовать поведение f при стремлении размера системы к бесконечности. Таким образом, Гионе и Фигалли очень красиво доказывают универсальность для широкого класса систем типа собственных значений случайных матриц.

Из других наград Международного математического союза я бы выделил медаль Черна, вручённую на конгрессе *Масаки Кашиваре*. Профессор Кашивара хорошо знаком российским математикам, и его работы по теории D-модулей, соответствию Римана — Гильберта, теории Каждана —

Люстига, квантовым группам, кристальным базисам и т. д. очень ценят, читают и развивают в нашей стране. Уверен, что очень многие присоединятся к моим самым сердечным поздравлениям Кашиваре с получением медали Черна.

Разумеется, получить награду ММС — это большая честь, и все лауреаты заслуживают самых тёплых поздравлений. Особенно эмоциональным получилось вручение премии Лилавати, которой в 2018 году был удостоен турецкий математик *Али Несин*. Премия Лилавати вручается с 2010 года и отмечает выдающиеся вклады в популяризацию математики и в укрепление роли математики в обществе.

После аспирантуры Йельского университета Али Несин преподавал математику в университетах США, но вернулся в Турцию после смерти своего отца, известного турецкого писателя и драматурга Азиза Несина (1915–1995). Азиз Несин основал в 1973 году специальный фонд с целью помочь получить образование тем детям, которые лишены такой возможности. Али возглавил работу этого фонда, преподавал математику в одном из университетов Стамбула и был вовлечён во множество проектов, самым заметным из которых было создание «Математической деревни» в одном из удалённых уголков Эгейского побережья Турции. Хотя эта деревня построена на чистом энтузиазме и в бытовом смысле довольно аскетична, она стала местом проведения летних математических лагерей для школьников, конференций и других мероприятий. Успех этой деревни, а также гражданское мужество, проявленное организаторами при её создании, были особенно отмечены в laudatio.

Думается, что и в философии математического образования, и в преданности своему делу, и в плане духа подвижничества, у Али Несина есть огромное количество единомышленников и ничуть не менее опытных коллег в нашей стране. Невозможно очертить парой фраз весь российский опыт организации летних школ для школьников, но как не упомянуть знаменитую дубнинскую «Современную математику», которая соберётся этим летом уже в 19-й раз! Теперь уже, увы, без своего главного вдохновителя Виталия Арнольда. В умении довольствоваться малым и при этом создавать всемирно известные математические центры нам тоже опыта не занимать, взять хотя бы знаменитый дачный семинар, организуемый Валерием Лунцем. Поэтому, поздравляя Али Несина и его соратника Севана Нишаняна с этой замечательной наградой, я также обращаю слова глубочайшей благодарности и признательности всем тем, кто вложил столь же большую часть своей жизни в дело математического образования и математического просвещения в нашей стране.

Из пленарных докладов, не связанных (увы!) с наградами, мне особенно запомнились доклады Джорджа Вильямсона и Рахула Пандхарипанде. Начнём с Вильямсона. Краеугольным фактом теории конечномерных комплексных представлений полупростых групп и алгебр Ли является установленная Германом Вейлем полупростота: каждое представление есть прямая сумма неприводимых представлений (которые, в свою очередь, описываются своими старшими весами и характеры которых даются элегантной формулой того же Германа Вейля). В бесконечномерной ситуации категория модулей со старшим весом над полупростой алгеброй Ли уже не является полупростой и характеры неприводимых представлений устроены гораздо хитрее. Изучение этой важной категории, начатое в работах И. Н. Бернштейна, И. М. Гельфанда и С. И. Гельфанда, вышло на новую орбиту в работах Каждана и Люстига. Каждан и Люстиг предложили гипотетическую формулу для разложения простейших, так называемых стандартных модулей на неприводимые, а тем самым и формулу для характеров неприводимых модулей. Ингредиенты в гипотезе Каждана — Люстига могут быть описаны как геометрически, в терминах особенностей многообразий Шуберта, так и комбинаторно, в терминах алгебры Гекке группы Вейля алгебры Ли. Доказательство гипотезы КЛ, вскоре найденное Бейлинсоном и Бернштейном и, независимо, Брылинским и Кашиварой следует, без сомнения, отнести к важнейшим достижениям теории представлений второй половины XX века. Дальнейшее переосмысление этого круга вопросов было достигнуто Бейлинсоном, Гинзбургом, Зёргелем (учеником которого был Вильямсон) и другими.

Большинство работ Вильямсона посвящено аналогичным вопросам для представлений над полем характеристики p. Хорошо известно, что теория представлений в характеристике p напоминает бесконечномерную теорию представлений в характеристике 0, в которой дополнительно возникает некоторая периодичность по модулю p. В частности, вместо зеркал $\alpha(x) = 0$, где α пробегает множество корней группы Вейля, в характеристике p следует рассматривать зеркала $\alpha(x) \in p\mathbb{Z}$, отражения в которых порождают аффинную группу Вейля (и соответствующую аффинную алгебру Гекке). Гипотетический аналог формулы для характеров неприводимых представлений был в такой постановке предложен Люстигом и даже доказан для всех очень больших p в работе Андерсена, Янцена и Зёргеля (с использованием важных результатов Кашивары и Танисаки и других). Более прямые доказательства, опять же для очень больших p, были предложены Безрукавниковым и его соавторами (Архиповым и Гинзбургом, в одном варианте, и Мирковичем и Рюминым — в другом).

Вильямсон добился замечательного прогресса в вопросе о том, когда и как формулы типа гипотезы Люстига модифицируются при не очень большом p. Как и многие исследователи до него, он строит геометрические модели кратностей с помощью конструктивных пучков и теорем типа теоремы о разложении Бейлинсона, Бернштейна и Делиня. Сложность в том, что теорема ББД не верна с коэффициентами в поле характеристики p, но Вильямсон и его соавторы придумали некоторую правильную её замену с использованием введённых ими так называемых пучков чётности. Вычисления с этими пучками чётности хоть и гораздо сложнее, чем вычисления в классической теории Каждана — Люстига, но всё же не безнадёжны и доводимы до ответа. В частности, вычисления Вильямсона показывают, что простые числа p, для которых гипотеза Люстига не выполняется для $\mathrm{SL}(n)$, могут расти экспоненциально с n. Это, конечно, совершенно поразило воображение всех специалистов, которые ожидали условия типа $p > n$ для справедливости формулы Люстига.

По моему скромному суждению, вся эта область математики наполнена необычайной красотой, и пожалуй, заслуги перечисленных мной математиков могли бы быть отмечены бóльшим числом наград Международного математического союза, чем одна медаль Черна, вручённая Кашиваре. Хотя, конечно, всегда полезно помнить, что подлинная цель занятий математикой лежит гораздо выше и что она же и есть наша главная награда.

Главной темой в докладе моего старого друга Рахула Пандхарипанде было пространство модулей кривых рода g, о которых мы уже говорили ранее. Как пространство модулей гладких кривых рода g, так и его компактификация стабильными кривыми, предложенная Делинем и Мамфордом, играют исключительно важную роль в алгебраической геометрии и математической физике. Это как бы старший, нелинейный брат многообразий Грассмана и других многообразий, параметризующих объекты линейной алгебры. Подобно тому как многообразия Грассмана и их когомологии объясняют геометрию векторных расслоений (иными словами, геометрию семейств векторных пространств), пространства модулей кривых объясняют геометрию семейств алгебраических кривых. В частности, они нужны повсюду, где изучается исчислительная геометрия кривых, т. е. при ответе на каждый вопрос типа: сколько кривых заданной степени и заданного рода в каком-то алгебраическом многообразии X удовлетворяют тем или иным условиям (например, пересекают заданные циклы в X). Исчислительные вопросы такого рода составляют современный аналог исчисления Шуберта. Они возникают не только в алгебраической геометрии, но и в математической физике, в частности в математических аспектах теории струн.

Полные кольца когомологий пространств модулей кривых устроены очень сложно. К счастью, в исчислительных задачах можно ограничиться только их малой частью, порождённой понятными геометрическими классами. Это так называемое тавтологическое кольцо, и поскольку образующие этого кольца можно предъявить явно, главный вопрос о тавтологическом кольце — это вопрос о соотношениях между этими образующими. Долгое время эта область находилась под влиянием гипотез, высказанных Карелом Фабером и утверждавших, среди прочего, что тавтологическое кольцо горенштейново, подобно кольцу когомологий гладкого полного многообразия некоторой размерности (а именно размерности $g-2$ для пространства модулей $\mathscr{M}_g$ гладких кривых рода g). Это было проверено для $g \leqslant 23$, но, видимо, неверно начиная с рода $g=24$. Дело в том, что в последние годы Пандхарипанде, его учеником Пикстоном и их соавторами, был достигнут замечательный прогресс в доказательстве и анализе соотношений в тавтологическом кольце. Всё указывает на то, что все соотношения уже найдены и это есть окончательный, негоренштейнов ответ. Не удивлюсь, кстати, если окажется, что $g = 24$ здесь связано с решёткой Лича, о которой пойдёт речь ниже.

Стоит сказать пару слов об одном важном ингредиенте анализа когомологий пространств модулей кривых. Обозначим через $\overline{\mathscr{M}}_{g,n}$ пространство модулей стабильных кривых рода g с n отмеченными точками. Напомним, что стабильным кривым разрешается иметь простые двойные особые точки вида $xy = 0$, если только они не отмечены. Рассоединяя две ветви в такой особой точке, мы получаем отображения

$$\overline{\mathscr{M}}_{g_1,n_1+1} \times \overline{\mathscr{M}}_{g_2,n_2+1} \to \overline{\mathscr{M}}_{g_1+g_2,n_1+n_2} \quad \text{и} \quad \overline{\mathscr{M}}_{g,n+2} \to \overline{\mathscr{M}}_{g+1,n}\,.$$

Набор классов когомологий $\overline{\mathscr{M}}_{g,n}$ для всех g и n называется когомологической теорией поля (название, пришедшее из математической физики), если он согласован с этими отображениями. Замечательным фактом о такого рода объектах является гипотеза А. Гивенталя, доказанная К. Телеманом. Она утверждает, что все такие теории, удовлетворяющие некоторому условию невырожденности, можно некоторым калибровочным преобразованием перевести в тривиальный набор из единичных классов когомологий. Эта теорема упаковывает всю априори неограниченную сложность многих когомологических теорий поля в одну матрицу, зависящую от параметра, что делает возможными многие ранее неприступные вычисления.

Из докладов на конгрессе нового поколения математиков не могу оставить без внимания доклад Е. Малинниковой и А. Логунова, а также доклад Марины Вязовской. Замечу, что все они недавно были удостоены

престижной премии Клэя и мне бы очень хотелось видеть и Марину, и Сашу в числе лауреатов новых премий ММС в 2022 году. Женя Малинникова уже не может получить Филдсовскую медаль по возрасту, но несомненно, что широкое признание её заслуг ждёт её в какой-то другой форме. Всех, кто читает по-английски, я призываю прочитать статью Генри Кона о Марине Вязовской в Известиях Американского математического общества[2)]. В ней рассказывается, как Кон и Элкис много лет тому назад придумали стратегию для доказательства оптимальности упаковок шаров, отвечающих решёткам E_8 и решётке Лича в размерностях 8 и 24 соответственно. И никто не мог подобрать к этой стратегии один магический ключ, одну магическую функцию до того, как Марина предъявила подобную функцию явно. Саша и Женя аналогично прославились простым красивым решением очень старой и хорошо известной задачи, а именно задачи об оценке меры множества нулей собственной функции оператора Лапласа.

К слову, о следующем математическом конгрессе: он пройдёт в нашей стране в 2022 году, в Санкт-Петербурге. Такое решение было принято на Генеральной ассамблее Международного математического союза, которая собиралась в бразильском городе Сан-Паулу прямо перед началом конгресса в Рио. Автор этих строк был там в составе делегации, представлявшей заявку Санкт-Петербурга. Наша заявка опередила при голосовании заявку, поданную городом Парижем, в очень сложной и драматичной борьбе, о которой можно почитать в репортаже Натальи Дёминой, написанном по свежим следам событий[3)].

Конечно, подобное международное признание заслуг отечественной математической школы не может не окрылять. (Заметьте, сколько имён петербургских математиков мы упомянули!) Но также следует помнить, что организация и успешное проведение конгресса — это огромный труд и огромная ответственность, тем более что обещали мы его провести на самом высоком уровне.

При всех успехах организаторов конгресса в Рио очень хотелось бы, чтобы у нас некоторые вещи получились лучше. К примеру, участие в конгрессе собственно бразильских и вообще южноамериканских математиков не было столь массовым, как хотелось бы. Очень хочется надеяться, что конгресс в Санкт-Петербурге будет очень притягательным событием для математиков из России и всех соседних государств.

[2)] *Cohn H.* A conceptual breakthrough in sphere packing // Notices Amer. Math. Soc. 2017. Vol. 64, № 2. P. 102–115.

[3)] https://trv-science.ru/2018/07/29/mezhdunarodnyj-kongress-matematikov-projdet-v-2022-v-sankt-peterburge/

Традиционно программу конгресса дополняют сателлитные конференции по множеству более специальных тем, и мы надеемся, что организация подобных мероприятий будет замечательным поводом укрепить дружбу и сотрудничество между российскими математиками, нашими соседями и всем мировым математическим сообществом.

Мы, организаторы конгресса в Санкт-Петербурге, очень надеемся, что студенты и аспиранты математических специальностей откликнутся на призыв стать волонтёрами конгресса. С одной стороны, присутствие на конгрессе в качестве волонтёра даёт уникальный шанс окунуться в самую гущу современной математики. С другой стороны, успех и само проведение конгресса невозможно без труда множества энтузиастов математики, в том числе, разумеется, организаторов конгресса и его волонтёров.

Отдельно хочется обратиться к читающим по-русски математикам за рубежом, нашим коллегам, которые или причисляют себя к воспитанникам некогда единой математической школы, или как-то по-иному с ней связаны. Очень ждём вас всех в Санкт-Петербурге! Хотя пленарные и секционные доклады конгресса всегда делаются на английском языке, мы планируем включить в программу ряд мероприятий на русском языке, в том числе некоторые из публичных лекций.

Хочется обратить внимание всех молодых математиков и всех математиков из развивающихся стран на то, что положения заявки предполагают полную поддержку для 1000 участников из развивающихся стран и оплату расходов в Санкт-Петербурге для 1300 молодых математиков. Эта поддержка будет предоставляться в соответствии с рекомендациями Международного математического союза и в партнёрстве с зарубежными математическими обществами и агентствами, финансирующими науку.

Среди важных решений Генеральной ассамблеи в Сан-Паулу было создание специального структурного комитета ММС. В задачу этого комитета входит разработка структуры программы конгрессов, в то время как программный комитет, который раньше полностью отвечал за программу, будет теперь подбирать конкретных докладчиков под заданную структурным комитетом матрицу. Конечно, мы с большим интересом ждём решений и рекомендаций структурного комитета, тем более что конгресс в Санкт-Петербурге будет первой пробой нового механизма составления программы. Всех, у кого есть на этот счёт соображения и пожелания, призываю обращаться прямо в структурный комитет через его председателя Теренса Тао.

Что касается публичных лекций, выставок, фестивалей, культурных и прочих мероприятий за рамками полномочий структурного и программного комитетов, то за это отвечает организационный комитет, и мы с радостью услышим любые пожелания и идеи по поводу этой части программы.

Наконец, в связи с проведением Международного конгресса математиков ожидается, что 2022 год будет объявлен в Российской Федерации годом математики. Мне видится, что год математики в стране мог бы быть наполнен массой мероприятий для детей, школьников, студентов, просто любителей математики, мероприятий очень разных как по теме, формату, так и по месту проведения. Что-то будет происходить в школах и других учебных заведениях, что-то в различных математических центрах, на страницах печати, на телевидении, в интернете и т. д. Было бы замечательно дать простор и дополнительный импульс тому интересу и склонности к математике, которые определённо есть у многих жителей нашей страны. Для масштабного и успешного проведения года математики потребуется участие не только значительной части математиков, работающих в нашей стране, но и нашей математической диаспоры. Я очень надеюсь, что читатели «Математического просвещения» примут самое активное участие в подготовке и проведении столь масштабного мероприятия.

Андрей Юрьевич Окуньков, Колумбийский университет
(Нью-Йорк, США), Сколтех (Москва), НИУ ВШЭ
okounkov@math.columbia.edu

Геометрия: классика и современность

Доказательство гипотезы Пуанкаре (по работам Г. Перельмана)

Л. Бессьер, Ж. Бессон, М. Буало

Гипотеза, изначально сформулированная как чисто топологическая, сопротивлялась атакам топологов в течение ста лет, чтобы сдаться геометрам. Программа исследований, начатая Ричардом Гамильтоном в 1982 году и завершённая Григорием Перельманом в 2003 году, основывается на понятии потока кривизны Риччи — уравнения эволюции, которое стремится сделать метрику однородной.

§ 1. Введение

Топология поверхностей была хорошо изучена уже к концу XIX века. Всякая ориентируемая поверхность без края может быть описана топологически как граница некоторого кренделя. Количество дырок в этом кренделе равно максимальному числу непересекающихся замкнутых кривых, которые можно провести на этой поверхности так, чтобы в результате она не распалась на части; это число, называемое родом, даёт полную классификацию поверхностей. Таким образом, самая простая поверхность с топологической точки зрения — это сфера $S^2 \subset \mathbb{R}^3$; она является границей шара радиуса 1, и любая проведённая на ней замкнутая кривая разбивает её на части.

Bessières L., Besson G., Boileau M. La preuve de la conjecture de Poincaré d'après G. Perelman // Images des mathématiques. CNRS. Le 15 octobre 2006.

Перевод с французского Е. Ю. Смирнова.

Аналогичное исследование гиперповерхностей (или многообразий) без края в более высоких размерностях было начато несколько позже, а именно в 1895 году, в заметке Анри Пуанкаре *Analysis Situs* [13], которая ознаменовала рождение современной алгебраической топологии.

В 1904 году в пятом и последнем дополнении к *Analysis Situs* [13] Пуанкаре построил пример, показывающий, что в размерности 3 единичная сфера $S^3 \subset \mathbb{R}^4$ не может быть охарактеризована тем свойством, что всякая вложенная в неё поверхность разделяет её на части; для этого требуется задействовать более тонкие топологические понятия. Чтобы отличить построенное им трёхмерное пространство от сферы S^3, Пуанкаре использует понятие *фундаментальной группы*. Это алгебраический инвариант, который был введён ещё в первой заметке *Analysis Situs*; в его определении участвуют замкнутые пути (петли) в данном пространстве, которые не могут быть стянуты в точку посредством какой-либо непрерывной деформации (такие петли называются *существенными*). Если существенных петель нет, то фундаментальная группа тривиальна; в таких случаях говорят, что пространство *односвязно*[1)]. Таковы, например, все сферы S^n размерности $n \geqslant 2$. Пуанкаре показывает, что построенное им пространство не односвязно.

В конце своей статьи он формулирует следующий вопрос[2)], впоследствии ставший знаменитым: «Возможно ли, чтобы фундаментальная группа трёхмерного многообразия V была тривиальной и при этом многообразие V не являлось сферой?»

Предположение о том, что всякое односвязное многообразие размерности 3 есть сфера S^3, известно как *гипотеза Пуанкаре*. Доказательство этой гипотезы оказалось крайне сложной задачей. Она допускает естественное обобщение — *гипотезу о геометризации*, сформулированную Уильямом Тёрстоном в 1970-х годах для описания всех многообразий размерности 3.

Тёрстон [15] предположил, что восьми однородных геометрий достаточно для описания элементарных «кирпичиков», из которых строятся все многообразия размерности 3 (см. также [14]). При этом в центр исследований многообразий размерности 3 гипотеза геометризации поставила методы дифференциальной геометрии.

В начале восьмидесятых годов Ричард Гамильтон сформулировал новую программу исследований для доказательства гипотезы геометризации и,

[1)] Это современная терминология. Пуанкаре использовал термин «односвязное» при обозначении сферы.

[2)] Здесь Пуанкаре использует термин *односвязное пространство*.

в частности, гипотезы Пуанкаре. Его подход был основан на понятии потоков Риччи: в пространстве римановых метрик на рассматриваемом многообразии исследуется поведение решений некоторого дифференциального уравнения, связанного с кривизной (поток Риччи). При эволюции потока Риччи метрика стремится к однородной, однако же распределение кривизны не является однородным: за конечное время кривизна может «накапливаться» и становиться бесконечной в некоторых точках многообразия. С этими явлениями, которые называются особенностями потока Риччи, Гамильтону не удалось полностью разобраться.

Недавно[3)] Григорий Перельман определил для потока Риччи монотонную функцию, называемую *энтропией*, которая позволила ему описать, каким образом возникают эти особенности, и классифицировать их. При этом ему удалось сделать решающий шаг в завершении программы Гамильтона. Начав с потока Риччи, он построил *поток с хирургией*, позволяющий избавиться от особенностей. Далее мы расскажем о доказательстве гипотезы Пуанкаре, которое было предложено Перельманом.

§ 2. Поток, ассоциированный с кривизной Риччи

Будем искать процесс типа эволюции, который приводил бы к некоторой выделенной римановой метрике на данном дифференцируемом многообразии M (см. определение 1). Мы хотим, чтобы у этой метрики была постоянная кривизна Риччи; такие метрики называются эйнштейновыми.

Поток, ассоциированный с кривизной Риччи (см. определение 1), — это дифференциальное уравнение на (бесконечномерном) пространстве $\mathscr{M}$ римановых метрик на многообразии M. В идеале хотелось бы, чтобы это обыкновенное дифференциальное уравнение использовало градиент некоторой функции, взятый со знаком «минус» (чтобы его траектории сходились к минимумам). Естественным кандидатом на эту роль является функция, которую в физике называют функционалом Гильберта — Эйнштейна. Это интеграл от кривизны (скалярная кривизна); критическими точками этой функции являются эйнштейновы метрики, т. е. такие метрики, для которых[4)] $\mathrm{Ric}_g = \lambda g$. К сожалению, простое вычисление показывает, что градиент этой функции даёт уравнение, которое в общем случае не имеет решений. Однако оказывается, что для наших целей годится некоторая модификация этого уравнения. Назовём потоком

[3)] Статья опубликована в 2006 г. — *Прим. перев.*

[4)] Обозначения см. на с. 57. — *Прим. перев.*

Риччи семейство $g(t)$ римановых метрик на M, определённое на $[0;T)$ и удовлетворяющее следующему уравнению эволюции (см. пример 2):

$$\frac{\partial g}{\partial t} = -2\,\mathrm{Ric}_{g(t)}. \tag{1}$$

Определение 1

Дифференцируемое многообразие размерности n — это пространство, которое локально выглядит как стандартное евклидово пространство $\mathbb{R}^n$: у каждой точки имеется открытая окрестность, гомеоморфная $\mathbb{R}^n$, которая называется *картой*, причём переход от одной карты к другой задаётся диффеоморфизмом класса C^∞ (см. рис. 1). В размерности 3 дифференцируемая структура единственна с точностью до диффеоморфизма, тогда как пространство $\mathbb{R}^4$ можно снабдить бесконечным числом попарно не диффеоморфных дифференцируемых структур. В дальнейшем все рассматриваемые многообразия мы будем считать принадлежащими классу C^∞ и ориентируемыми.

Чтобы изучать геометрию многообразия M (например, вычислять длины кривых, расстояния, объёмы и т. д.), требуется допол-

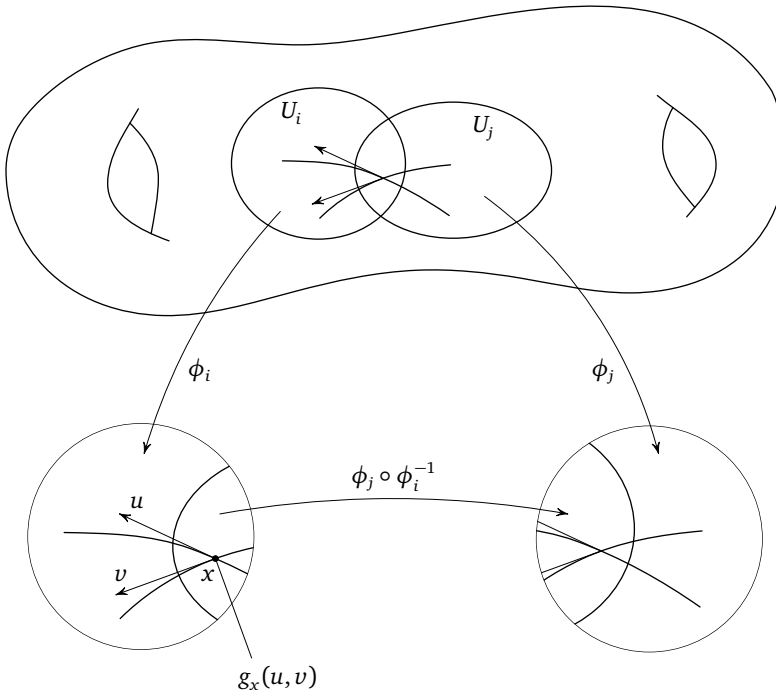


Рис. 1

нительная структура, называемая **римановой метрикой** и обозначаемая через g: в каждой точке x карты она задаёт скалярное произведение $g_i(x)$, которое изменяется от точки к точке как функция класса C^∞ и согласовано с переходом от одной карты к другой. Кривизны, ассоциированные с метрикой, измеряют её инфинитезимальное отклонение от стандартной метрики на $\mathbb{R}^n$. Они вычисляются как полиномиальные выражения от коэффициентов метрики g, ∂g и $\partial^2 g$.

Например, с каждой (двумерной) плоскостью P в касательном пространстве T_xM можно связать кривизну сечения $K(P)$, определяемую следующим образом. Обозначим через $C(r)$ круг с центром в x и радиусом r, касающийся плоскости P. Длина соответствующей окружности удовлетворяет формуле

$$\ell(C(r)) = 2\pi r\left(1 - \frac{K(P)}{6}r^2 + o(r^2)\right),$$

где $K(P)$ измеряет дефект, т. е. отклонение от евклидовой длины окружности. Кривизна Риччи (называемая также тензором Риччи) задаётся в каждой точке x пространства M симметрической билинейной формой на T_xM (не обязательно положительно определённой). Её значение на векторе $v \in T_xM$, обозначаемое $\mathrm{Ric}_g(v, v)_x$, вычисляется как сумма кривизн сечений плоскостями, порождёнными векторами v и e_i, где e_i пробегает ортонормированный базис ортогонального дополнения к v в T_xM. Она характеризует дефект площади сфер малого радиуса. **Скалярная кривизна** $R(x)$ — это функция на M, определённая в каждой точке x как след кривизны Риччи $\mathrm{Ric}_g(\cdot, \cdot)_x$ относительно скалярного произведения g_x, т. е. как сумма собственных значений соответствующего оператора. Она характеризует отклонение объёма шаров малого радиуса от евклидова объёма.

Подход с применением дифференциальных уравнений, описанный выше, оказывается труднопримени́мым на бесконечномерном пространстве $\mathscr{M}$. Вместо этого мы применим более эффективный подход, при котором уравнение (1) записывается в некоторой системе локальных координат, параметризующих $\mathscr{M}$. При этом (1) становится параболическим уравнением в частных производных типа *«реакция-диффузия»*.

Пример 2. На сфере решением потока Риччи является метрика $g(t) = (1 - 2\lambda t)g_0$ на промежутке $[0; 1/(2\lambda))$, если кривизна Риччи в исходной метрике равна $\mathrm{Ric}_{g_0} = \lambda g_0$ при $\lambda > 0$.

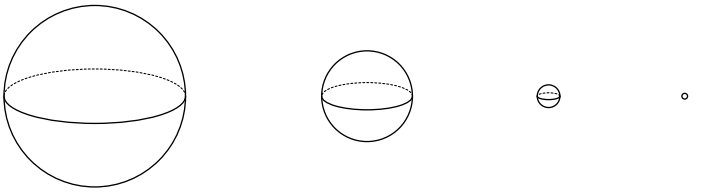


Рис. 2

Более общим образом, эволюция потока задаётся при помощи гомотетий, если начальная метрика является эйнштейновой. Метрики отрицательной кривизны при эволюции растягиваются, а метрики положительной кривизны стягиваются (см. рис. 2).

§ 3. Работы Р. Гамильтона

Гамильтон доказал [5] существование решения на малом отрезке времени для произвольных гладких начальных условий (простое доказательство см. в [4]). Затем этот поток можно продолжить, так как кривизны сечения (см. определение 1) остаются ограниченными по абсолютной величине. Чтобы оценить значения кривизн, нужно записать их уравнения эволюции и использовать принцип максимума.

Принцип максимума для кривизн

Начнём со скалярной кривизны, которая эволюционирует в соответствии с параболическим уравнением

$$\frac{\partial R}{\partial t} = \Delta R + 2|\operatorname{Ric}|^2, \tag{2}$$

где все величины являются функциями от $g(t)$. Во всякой точке x минимума скалярной кривизны метрики $g(t)$ лапласиан ΔR неотрицателен, следовательно, $\partial R/\partial t \geqslant 0$. Можно выдвинуть естественное предположение, что минимум скалярной кривизны метрики $g(t)$ на многообразии M (обозначим его через $R_{\min}(t)$) возрастает при росте t. Принцип максимума позволяет доказать это строго. Более того, если $R_{\min}(0) > 0$, можно доказать, что $R_{\min}(t)$ стремится к $+\infty$ за конечное время (если такой поток существует). В этом случае мы убеждаемся, что максимум кривизн сечения стремится к $+\infty$ за конечное время. Чтобы получить дополнительную информацию, используем уравнение эволюции тензора Риччи, имеющее следующий вид:

$$\frac{\partial \operatorname{Ric}}{\partial t} = \Delta \operatorname{Ric} + Q(\operatorname{Ric}), \tag{3}$$

где Q — некоторое квадратичное выражение. Векторный принцип максимума показывает, что если $\mathrm{Ric}_{g_0} \geqslant 0$, то и $\mathrm{Ric}_{g(t)} \geqslant 0$. Если, кроме того, $\mathrm{Ric}_{g_0} > 0$, то это же верно и для всех t, причём во всех точках многообразия имеет место оценка

$$\frac{1}{R}\left|\mathrm{Ric} - \frac{R}{3}g\right| \leqslant \frac{\alpha}{R^{\beta}}, \tag{4}$$

где α и β — положительные константы. Это означает, что при стремлении $R(x,t)$ к бесконечности относительная разница в точке x между $\mathrm{Ric}_{g(t)}$ и его средним значением $\frac{R}{3}g$ стремится к нулю. При помощи оценки градиента скалярной кривизны Гамильтон доказал, что в предположении о строгой положительности кривизны Риччи она стремится на конечном промежутке времени к бесконечности с одной и той же скоростью во всех точках. Итак, если перенормировать метрику так, чтобы объём был постоянным, то $g(t)$ будет сходиться к некоторой метрике с постоянной положительной кривизной сечений, откуда получается

Теорема 3. *Если M — замкнутое риманово многообразие, на котором можно ввести метрику со строго положительной кривизной Риччи, то на M также существует метрика с постоянной и положительной кривизной сечений.*

Замечание 4. В частности, M является факторпространством сферы S^3 по конечной группе изометрий. Такое многообразие называется **сферическим**. Например, таково пространство прямых в $\mathbb{R}^4$, являющееся факторпространством сферы S^3 по отождествлению противоположных точек; при этом получается проективное пространство, обозначаемое через $\mathbb{P}^3(\mathbb{R})$. Это основополагающая теорема во всей этой теории; она является первым шагом к доказательству гипотезы Пуанкаре.

Ситуация резко меняется, если кривизна Риччи не является строго положительной. Наиболее общий результат таков: *для произвольных начальных условий поток существует на некотором максимальном промежутке $[0;T)$, и если $T<\infty$, то максимум кривизн сечений в момент времени t стремится к $+\infty$ при $t \to T$.* В последнем случае момент времени T называют *особым*. Вообще говоря, кривизна неограниченно возрастает лишь на некоторой части многообразия; при этом говорят, что поток имеет особенность. Между тем вариант предыдущих результатов, *теорема о сжатии* Гамильтона — Айви, показывает, что отрицательная часть кривизны становится пренебрежимо малой по сравнению со скалярной кривизной. Таким образом, скалярная кривизна регулирует значения всех кривизн.

Теперь скажем несколько слов об изучении особенностей.

Изучение особенностей: техника масштабирования

Эта техника, в анализе являющаяся классической, была использована в данной ситуации Гамильтоном в работах [7] и [8]. Масштабирование заключается в растяжении метрики и замедлении течения времени, которые позволяют получить новое решение уравнения потока. Будем рассматривать последовательности масштабирований и переходить по ним к пределу (см. ниже). Если доказать существование предельных потоков и классифицировать их, это даст нам всевозможные модели особенностей. Вопрос о существовании предельных потоков был одним из основных препятствий на пути к осуществлению программы Гамильтона. Этот вопрос недавно[5)] был полностью решён Перельманом.

В общем случае рассматривается последовательность масштабирований в точках (x_k, t_k), для которых $Q_k := R(x_k, t_k) \to +\infty$ и скалярная кривизна на $M \times [0; t_k]$ в которых максимальна. Тогда последовательность параболических растяжений метрик $g_k(t)$ в (x_k, t_k) имеет ограниченную кривизну на отрезках $[-t_k Q_k; 0]$, сходящихся к $(-\infty; 0]$. Подходящим условием, чтобы гарантировать сходимость последовательности $(M, g_k(t), x_k)$ (или некоторой её подпоследовательности) — в некотором смысле, который мы здесь не уточняем, — к потоку $(M_\infty, g_\infty(t), x_\infty)$, является оценка снизу на объём единичного шара с центром в x_k (в метрике $g_k(0)$) некоторой положительной константой, не зависящей от k. Первый из замечательных результатов Перельмана [10] состоит в том, что эта оценка всегда выполнена, если кривизна стремится к бесконечности за конечное время. По построению, полученный предел является потоком на $(-\infty; 0]$ ограниченной и при этом ненулевой кривизны. Более того, теорема Гамильтона — Айви о сжатии позволяет показать, что кривизна сечений предельного потока положительна или равна нулю.

Параболические растяжения. Идея масштабирования формализуется при помощи понятия **параболического растяжения** (см. рис. 3). Если даны поток Риччи $g(t)$ на $M \times [0; T)$, точка x_0 и момент времени t_0, то параболическое растяжение определяется как решение потока, заданное формулой

$$g_0(t) = Q_0 \cdot g\Big(t_0 + \frac{t}{Q_0}\Big), \quad \text{где } Q_0 = R(x_0, t_0).$$

Оно определено на промежутке $[-t_0 Q_0; (T - t_0)Q_0)$ и нормируется условием $R_{g_0}(x_0, 0) = 1$.

[5)] См. примечание 3.

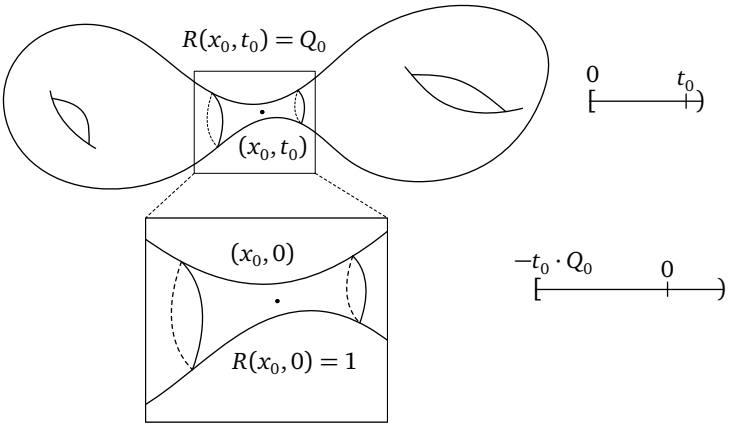


Рис. 3

§ 4. Потоки с хирургией по Перельману

Основным результатом первой статьи Перельмана [10] является теорема о канонических окрестностях, описывающая метрику $g(t)$ в точках большой скалярной кривизны. Если кривизна велика повсеместно, из этого получается классификация многообразий M. В таких случаях говорят, что поток останавливается. В противном случае идея (восходящая к Гамильтону [9]) состоит в том, чтобы избавиться от кусков многообразия M с большой кривизной, разрезав его вдоль сфер S^2 и заклеив дыры шарами B^3. Разумеется, это следует делать, следя за топологией и геометрией многообразия. После этого нужно рассмотреть поток на новом многообразии, которое, возможно, не будет связным, и повторить описанную процедуру. В некоторых случаях отдельные связные компоненты исчезают при хирургических операциях. В работе [11] Перельман доказывает, что для произвольных начальных условий, нормированных должным образом, этот поток с хирургией можно продолжать бесконечно. На каждом конечном промежутке времени при этом делается лишь конечное число хирургических операций. Если весь поток останавливается за конечное время, можно описать все его связные компоненты, а следовательно, и исходное многообразие. Так осуществляется доказательство гипотезы Пуанкаре. Классификация для длинных отрезков времени более сложна, и здесь мы её не рассматриваем.

Канонические окрестности

Теорема о канонических окрестностях утверждает, что, грубо говоря, в точках с большой скалярной кривизной потока Риччи геометрия является

канонической, т. е. почти что изометричной одной из конечного числа простых моделей. Чтобы не вводить лишних параметров, предположим, что поток живёт на промежутке, содержащем отрезок $[0;1]$ (этого можно добиться перенормировкой начальной метрики), и потребуем, чтобы единичные шары в начальной метрике были почти евклидовыми. Такие начальные данные будем называть нормализованными. Имеет место

Теорема 5. *Для всякого достаточно малого $\varepsilon > 0$ существует универсальная константа $r = r(\varepsilon) > 0$ со следующим свойством. Пусть $(M, g(t))$ — поток Риччи с нормализованными начальными данными, а $x \in M$ и $t \geqslant 1$ таковы, что $R(x,t) \geqslant r^{-2}$. Тогда точка x имеет окрестность, которая после растяжения в $\sqrt{R(x,t)}$ раз становится изометричной с точностью до ε одной из следующих моделей*:

i) *цилиндру $S^2 \times (-1/\varepsilon; 1/\varepsilon)$ с канонической метрикой произведения и скалярной кривизной* 1; *такую окрестность назовём ε-горлышком*;

ii) *шару B^3 или дополнению к шару в проективном пространстве, т. е. $\mathbb{P}^3(\mathbb{R}) \setminus \bar{B}^3$, с метрикой строго положительной скалярной кривизны, которые близки всюду, кроме некоторого компакта, к описанному выше сферическому цилиндру*; *такую окрестность будем называть ε-шапочкой*;

iii) *замкнутому многообразию со строго положительной кривизной сечения*.

При этом будем говорить, что $g(t)$ удовлетворяет гипотезе о канонических окрестностях для масштаба r (см. рис. 4).

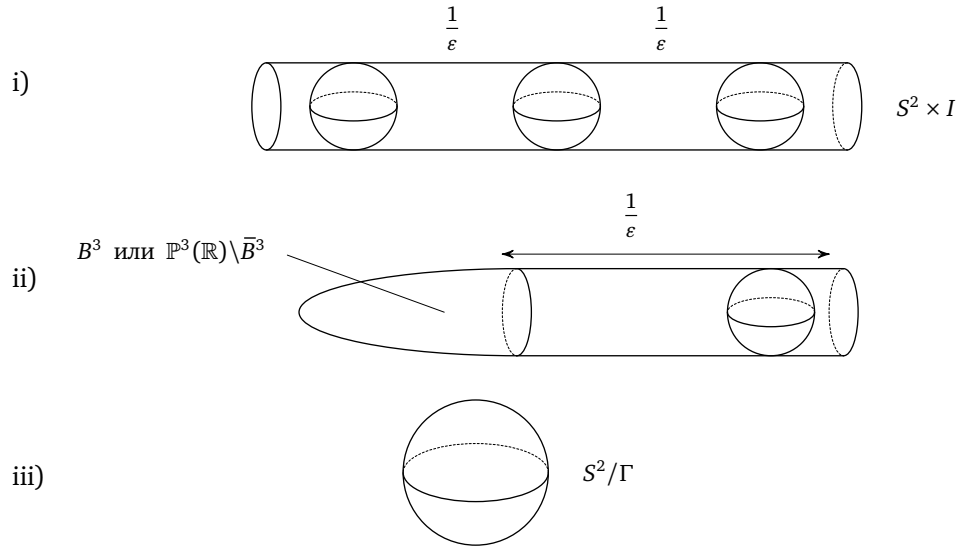


Рис. 4

Мы называем два диффеоморфных многообразия изометричными с точностью до ε, если не более чем на ε отличаются их римановы метрики, а также их производные порядков не более $1/\varepsilon$. В частности, кривизны на этих окрестностях сравнимы со скалярной кривизной $R(x,t)$. Размер окрестностей, соответствующих случаям i) и ii), сравним с $\frac{2}{\varepsilon}R^{-1/2}(x,t)$. Кроме того, пространственные и временны́е колебания оцениваются универсальными константами.

Замечание 6. В случае iii), ввиду связности, многообразие M целиком содержится в окрестности, и согласно теореме 3 оно диффеоморфно сферическому многообразию.

Описание первого особого момента времени

На протяжении этого раздела зафиксируем число $\varepsilon > 0$ и масштаб $r > 0$, относительно которого метрика $g(t)$ удовлетворяет предположению о канонических окрестностях. Мы будем описывать метрику $g(t)$ при $t \to T < \infty$, где T — особый момент времени. Обозначим через Ω множество точек, где скалярная кривизна остаётся ограниченной, т. е.

$$\Omega = \{x \in M, R(x,\cdot) \leqslant c(x) < +\infty\}.$$

По предположению, существует такая точка $x \in M$, в которой $R(x,t) \to +\infty$, следовательно, множество Ω строго меньше, чем M.

Множество Ω пусто: поток останавливается. В этом случае можно доказать, что M является сферическим многообразием, произведением $S^2 \times S^1$ или связной суммой проективных пространств, которую мы обозначаем через $\mathbb{P}^3(\mathbb{R})\#\mathbb{P}^3(\mathbb{R})$. Действительно, если кривизна всюду неограниченно возрастает, можно найти такой момент времени t_0, близкий к особому моменту T, что $(M, g(t_0))$ будет покрыто конечным числом канонических окрестностей. Если среди них имеется окрестность типа iii), то M диффеоморфно сферическому многообразию. В противном случае можно покрыть M горлышками, склеив их друг с другом по краям так, что в результате они образуют $S^2 \times S^1$, или же заклеив их края шапочками, получая в итоге S^3, $\mathbb{P}^3(\mathbb{R})$ или $\mathbb{P}^3(\mathbb{R})\#\mathbb{P}^3(\mathbb{R})$.

Замечание 7. Если M односвязно, то оно оказывается диффеоморфным сфере S^3.

Множество Ω непусто. С помощью оценок на кривизны можно доказать, что Ω — открытое множество, на котором метрика $g(t)$ сходится к регулярной метрике $g(T)$. Переходя к пределу, получаем, что $g(T)$ удовлетворяет предположению о канонических окрестностях. Чтобы понять

структуру множества Ω, зададимся масштабом кривизны $\rho < r$ и определим множество $\Omega_\rho = \{x \in \Omega\colon R(x,T) \leqslant \rho^{-2}\}$. Множество $\Omega \setminus \Omega_\rho$ покрывается горлышками и шапочками. Разбор различных случаев показывает, что любая точка $x \in \Omega \setminus \Omega_\rho$ принадлежит одному из следующих множеств:

i) *ε-трубка*: цилиндр $S^2 \times I$, полученный объединением конечного числа горлышек, граница которого содержится в Ω_ρ;

ii) *ε-остриё*: объединение бесконечного числа горлышек, диффеоморфное $S^2 \times \mathbb{R}^+$; конец острия $S^2 \times \{0\}$ принадлежит Ω_ρ, а на другом конце скалярная кривизна стремится к плюс бесконечности;

iii) объединение конечного числа горлышек, заклеенное шапочкой и примыкающее по границе к Ω_ρ;

iv) отдельные компоненты связности Ω_ρ: диффеоморфные $S^2 \times \mathbb{R}$ *двойные острия*, полученные как объединения бесконечного числа горлышек, и диффеоморфные $\mathbb{R}^3$ *заострённые шапочки*, полученные как объединение бесконечного числа горлышек, заклеенное шапочкой (см. рис. 5).

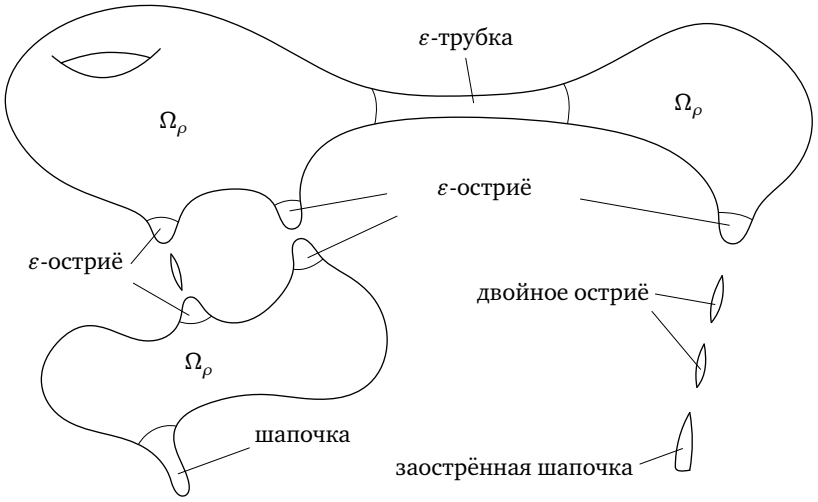


Рис. 5

Замечание 8. Если Ω_ρ пусто, то, действуя аналогично случаю $\Omega = \varnothing$, можно доказать, что M диффеоморфно $S^2 \times S^1$, $\mathbb{P}^3(\mathbb{R})$, $\mathbb{P}^3(\mathbb{R}) \# \mathbb{P}^3(\mathbb{R})$ или сферическому многообразию. В таких случаях тоже говорят, что поток останавливается, даже если кривизна не стремится к бесконечности во всех точках.

Хирургия. К множеству Ω можно применять следующие хирургические операции:

1° удаление всех компонент связности множества Ω, не пересекающих Ω_ρ;

2° выбрасывание множеств точек, примыкающих к Ω_ρ, и заклейка образовавшихся дыр шапочками, диффеоморфными шару (см. рис. 6).

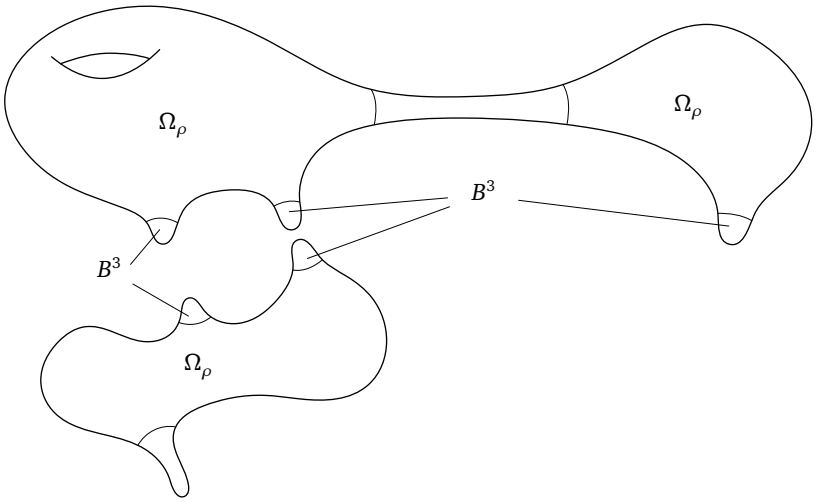


Рис. 6

В результате получается новое дифференцируемое многообразие, возможно несвязное, которое мы обозначаем через M_1. Во всякий момент времени $t < T$, близкий к T, множество $(M \setminus \Omega_\rho, g(t))$ покрыто каноническими окрестностями. Также можно проверить, что M есть сумма различных связных компонент многообразия M_1 и, возможно, конечного числа ручек $S^2 \times S^1$ и проективных пространств $\mathbb{P}^3(\mathbb{R})$.

Эта хирургическая операция должна выполняться с учётом метрики: нужно следить, как именно производятся все переклейки. Это делается так: нужно выбросить все точки в середине некоторого δ-горлышка, где $0 < \delta \ll \varepsilon$. Таким образом мы определяем хирургическую операцию с параметрами (r, δ), где r — это радиус кривизны (зависящий от ε), начиная с которого существуют канонические окрестности, и параметр ρ определяется как $\rho = \delta r$. На многообразии M_1, таким образом, появляется корректно определённая риманова метрика $g_1(T)$, которая задаёт начальные условия для уравнения (1), после чего поток можно рассматривать одновременно на всех компонентах связности многообразия M_1.

Замечание 9. Если множество Ω_ρ пусто, определение хирургической операции, тем не менее, имеет смысл. Но в этом случае $M_1 = \varnothing$, и поток останавливается.

Потоки с хирургией. Важное достижение Перельмана в работе [11] состоит в том, что ему удалось проитерировать описанную конструкцию бесконечное число раз. Зафиксируем $\varepsilon > 0$, которое используется в этой конструкции.

Определение 10. Пусть $r(t)$, $\delta(t)$ — пара функций, убывающих на луче $[0; +\infty)$ и принимающих положительные значения. Потоком с хирургией называется следующий набор данных:

i) строго возрастающая дискретная последовательность $(t_k)_{0 \leqslant k \leqslant N \leqslant \infty}$ со значениями из $[0; +\infty)$ и определённые для каждого целого числа k;

ii) компактное многообразие M_k, возможно несвязное или пустое;

iii) поток Риччи $g_k(t)$ на $M_k \times [t_k; t_{k+1})$, особый в точке t_{k+1} и удовлетворяющий гипотезе о канонических окрестностях для масштаба $r(t)$; при этом риманово многообразие $(M_{k+1}, g_{k+1}(t_{k+1}))$ получается из $(M_k, g_k(t))$ при помощи хирургической операции с параметрами (r, δ) в момент времени t_{k+1}.

Говорят, что риманово многообразие (M, g_0) нормализовано, если его кривизны сечений ограничены по модулю числом 1, а объём всякого единичного шара равен как минимум половине евклидова объёма. Перельман доказывает, что существуют такие универсальные строго убывающие функции $r(t)$, $\delta(t)$, для которых поток с хирургией существует на $[0; +\infty)$ для произвольных нормализованных начальных данных (M, g_0).

В частности, на каждом конечном отрезке совершается только конечное число операций. Если M_k — многообразие, полученное в k-й особый момент времени (с учётом выброшенных компонент связности), то M получается как связная сумма компонент связности многообразия M_k с некоторым количеством копий $S^2 \times S^1$ и факторов сферы S^3 по конечным группам. Если M_k пусто, поток останавливается, и M оказывается диффеоморфным связной сумме конечного числа многообразий вида $S^2 \times S^1$ и факторов S^3 по конечным группам. В частности, если M односвязно, то оно диффеоморфно S^3.

§ 5. Гипотеза Пуанкаре

После того как мы доказали существование потока с хирургией на бесконечном промежутке времени, для доказательства гипотезы Пуанкаре остаётся убедиться, что этот поток за конечное время будет останавливаться на гомотопической сфере, т. е. на односвязном многообразии. Согласно вышесказанному это многообразие тогда будет диффеоморфно S^3. Расскажем об этом более подробно.

Пусть у нас имеется односвязное компактное многообразие M_0, которое мы будем предполагать неприводимым (см. ниже). Возьмём на нём нормализованную метрику g_0. Для этих начальных данных можно построить поток с хирургией $(M_k, g_k(t))$, определённый на $[0; +\infty)$. Мы знаем, что у всякого многообразия M_k, если оно непусто, имеется компонента M_k^1, диффеоморфная M_0, а остальные компоненты являются сферами. Поэтому можно рассмотреть ограничение потока с хирургией на эту единственную компоненту. Чтобы доказать, что оно останавливается за конечное время, приведём набросок рассуждения Т. Колдинга и В. Миникоцци [3], которое технически более просто, чем доказательство Перельмана [12].

Неприводимость. Ориентируемое многообразие M называется неприводимым, если всякая сфера $S^2 \subset M$ ограничивает шар B^3. Из этого следует, что если M является связной суммой двух многообразий, то одно из них диффеоморфно самому M, а другое сфере S^3. Теорема Кнезера утверждает, что всякое ориентируемое многообразие является связной суммой конечного числа неприводимых односвязных многообразий и нескольких экземпляров $S^2 \times S^1$. В частности, если многообразие M односвязно, оно является связной суммой конечного числа неприводимых односвязных многообразий.

Ширина $(M_0, g(t))$ определяется как минимакс энергии сфер S^2 в «следе», оставленном многообразием M_0. Это геометрическая величина, строго положительная в случае, если петля, которая определяет след, существенна в пространстве $\mathscr{H}$, состоящем из непрерывных отображений из S^2 в $(M_0, g(t))$ с ограниченной энергией. Наличие существенной петли в $\mathscr{H}$ следует из односвязности многообразия M_0. Зафиксируем раз и навсегда гомотопический класс β некоторой существенной петли в $\mathscr{H}$. Ширина $W([\beta], g(t))$ риманова многообразия $(M_0, g(t))$ определяется по формуле

$$W([\beta], g(t)) = \inf_{\gamma \in [\beta]} \sup_{s \in [0;1]} E(\gamma(s)),$$

где

$$E(f) = \int_{S^2} |df|^2_{g(t)} d\mathrm{vol}_{S^2}$$

есть энергия отображения $f : S^2 \to (M_0, g(t))$.

Доказательство остановки потока за конечное время основано на следующих двух фактах.

1° На гладких частях потока ширина $W([\beta], g(t))$ достаточно быстро убывает при перемещении вдоль потока. Это следует из неравенства

Колдинга — Миникоцци [3]:

$$\frac{dW([\beta], g(t))}{dt} \leqslant -4\pi + \frac{3}{4(t+C)} W([\beta], g(t))$$

(C — некоторая константа, которую можно вычислить). Это гарантирует остановку за конечное время, если поток остаётся гладким, поскольку ширина становится нулевой за конечное время, но, с другой стороны, она должна быть положительной.

2° Если t_{k+1} — особый момент времени для потока $g_k(t)$ на M_0, то

$$\lim_{t \to t_{k+1}^-} W([\beta], g_k(t)) \geqslant W([\beta], g_{k+1}(t_{k+1})).$$

Отсюда следует существование $(1 + \xi(t))$-липшицева диффеоморфизма между $(M_0, g_k(t))$ и $(M_0, g_{k+1}(t_{k+1}))$, где $\xi(t) \to 0$ при $t \to t_{k+1}$.

Список литературы

[1] *Bessières L.* Conjecture de Poincaré: la preuve de R. Hamilton et G. Perelman // Gaz. Math. 2005. Vol. 106. P. 7–35.

[2] *Besson G.* Preuve de la conjecture de Poincaré en déformant la métrique par la courbure de Ricci, d'après G. Perelman // Séminaire Bourbaki 2004/2005, Astérisque. № 307 (2006). P. 309–347.

[3] *Colding T. H., Minicozzi W. P., II.* Estimates for the extinction time for the Ricci flow on certain 3-manifolds and a question of Perelman // J. Amer. Math. Soc. 2005. Vol. 18, № 3. P. 561–569.

[4] *Deturck D.* Deforming metrics in the direction of their Ricci tensors // J. Differential Geom. 1983. Vol. 18, № 1. P. 157–162.

[5] *Hamilton R.* Three-manifolds with positive Ricci curvature // J. Differential Geom. 1982. Vol. 17, № 2. P. 255–306.

[6] *Hamilton R.* Four-manifolds with positive curvature operator // J. Differential Geom. 1986. Vol. 24, № 2. P. 153–179.

[7] *Hamilton R.* The formation of singularities in the Ricci flow // Surveys in Differential Geometry, vol. II. (Cambridge, MA, 1993). Cambridge, MA: Int. Press, 1995. P. 7–136.

[8] *Hamilton R.* A compactness property for solutions of the Ricci flow // Amer. J. Math. 1995. Vol. 117, № 3. P. 545–572.

[9] *Hamilton R.* Four-manifolds with positive isotropic curvature // Comm. Anal. Geom. 1997. Vol. 5, № 1. P. 1–92.

[10] *Perelman G.* The entropy formula for the Ricci flow and its geometric applications. ArXiv:math.DG/0211159.

[11] *Perelman G.* Ricci flow with surgery on three-manifolds. ArXiv:math.DG/0303109.

[12] *Perelman G.* Finite extinction time for the solutions to the Ricci flow on certain three-manifolds. ArXiv:math.DG/0307245.

[13] *Poincaré H.* Œuvres. Tome VI. Gauthier-Villars. Paris, 1953.

[14] *Scott P.* The geometries of 3-manifolds // Bull. London Math. Soc. 1983. Vol. 15, № 5. P. 401–487. (Рус. пер.: *Скотт П.* Геометрии на трёхмерных многообразиях. М.: Мир, 1986.)

[15] *Thurston W. P.* Three dimensional manifolds, Kleinian groups and hyperbolic geometry // Bull. Amer. Math. Soc. (N. S.). 1982. Vol. 6, № 3. P. 357–381.

Лоран Бессьер, Гренобльский университет

Жерар Бессон, Гренобльский университет

Мишель Буало, Тулузский университет

Выпуклая и комбинаторная геометрия

О проблеме Крума

А. С. Безикович

От редакции

В 2001 г. на Московской математической олимпиаде была предложена следующая задача (11 класс, № 5):

Докажите, что в пространстве существует такое расположение 2001 *выпуклого многогранника, что никакие три из многогранников не имеют общих точек, а любые два касаются друг друга* (*т. е. имеют хотя бы одну граничную точку, но не имеют общих внутренних точек*). (*А. Канель*)

За решение этой задачи получил специальную премию имени Б. Н. Делоне ученик 11 класса школы № 57 г. Москвы Илья Межиров.

Вместо слова «касающиеся» мы будем далее использовать более точный термин «смежные». Понятно, что число 2001 в условии этой задачи несущественно и фактически требуется доказать

Утверждение 1. *Для любого n в пространстве существует n попарно смежных выпуклых многогранников.*

Вскоре после олимпиады в интернете была найдена статья [2], в которой доказывалось более сильное утверждение:

Для любого n в пространстве существует n попарно смежных выпуклых конгруэнтных многогранников.

С доказательством этого утверждения на русском языке можно ознакомиться в статье [4].

В статье А. С. Безиковича [1], перевод которой публикуется ниже, доказывается другое усиление утверждения 1:

В пространстве существует бесконечная последовательность выпуклых попарно смежных многогранников.

Отметим, что вместе с этой статьёй была опубликована статья Р. Радо [3], где доказывается более сложный и общий результат:

В кубе любой размерности n существует такая бесконечная последовательность выпуклых многогранников, что для любого натурального $k \leqslant \frac{1}{2}(n+1)$ любые k из этих многогранников имеют пересечение размерности $n-k+1$.

В этой статье[1)] приводится решение следующей проблемы М. Крума.

Каково максимально возможное число выпуклых многогранников, любые два из которых не имеют общих внутренних точек, но имеют общую границу положительной площади?

На плоскости ответ на аналогичный вопрос равен четырём, поэтому можно ожидать, что и для проблемы Крума ответом будет не очень большое число, например, 10 или 12. Однако, как будет показано, ответ равен бесконечности.

Введём в пространстве декартову систему координат, возьмём точку $A_0(1,0,0)$ и построим в вертикальной плоскости XOZ лежащую над OX выпуклую вниз ломаную $A_0A_1A_2\ldots$ с длиной, меньшей $1/4$, и такую, что угол между направлением оси OX и каждым из векторов $\overrightarrow{A_nA_{n+1}}$ больше $3\pi/4$.

Также возьмём такую последовательность[2)] положительных чисел δ_n, что $\sum_{n=1}^{\infty}\delta_n < 1/4$. Соединим точку $B_0(0,1,0)$ с A_0 и возьмём на отрезке B_0A_0 такую точку D_1, что $B_0D_1 = \delta_1$ (см. рис. 1); соединим D_1 с A_1 и возьмём на отрезке D_1A_1 такую точку D_2, что $D_1D_2 = \delta_2$; соединим D_2 с A_2 и возьмём на отрезке D_2A_2 такую точку D_3, что $D_2D_3 = \delta_3$, и т. д. Обозначим через r_0 плоскость XOY, а через r_n $(n = 1, 2, 3, \ldots)$ — плоскость, проходящую через точки A_{n-1}, D_n, A_n. Пусть r_n пересекает оси OY и OZ в точках B_n и C_n соответственно. Легко видеть, что все точки B_n лежат на положительной полуоси OY.

Обозначим через S_{k+1}, $k=0,1,2,\ldots$, многогранник, состоящий из всех точек, лежащих в положительном ортанте не ниже каждой из плоскостей $r_0,\ldots,r_k$ и не выше плоскости r_{k+1}.

Иными словами, если x^+, y^+ — полупространства $x \geqslant 0$, $y \geqslant 0$, а полупространства r_n^+, r_n^- состоят из всех точек, лежащих соответственно не ниже и не выше плоскости r_n, то $S_{k+1} = x^+ \cap y^+ \cap r_0^+ \cap \ldots \cap r_k^+ \cap r_{k+1}^-$. Будучи пересечением полупространств, многогранник S_{k+1} является выпуклым. Также легко видеть, что треугольник $D_{k+1}C_{k+1}A_{k+1}$ лежит на общей границе многогранников S_{k+1} и S_{k+2}, которые, таким образом, удовлетворяют

[1)] Перевод А. А. Заславского.

[2)] Ограничения на углы и $\{\delta_n\}$ можно было бы ослабить; их смысл в том, чтобы сделать возможной последующую конструкцию. — *Прим. перев.*

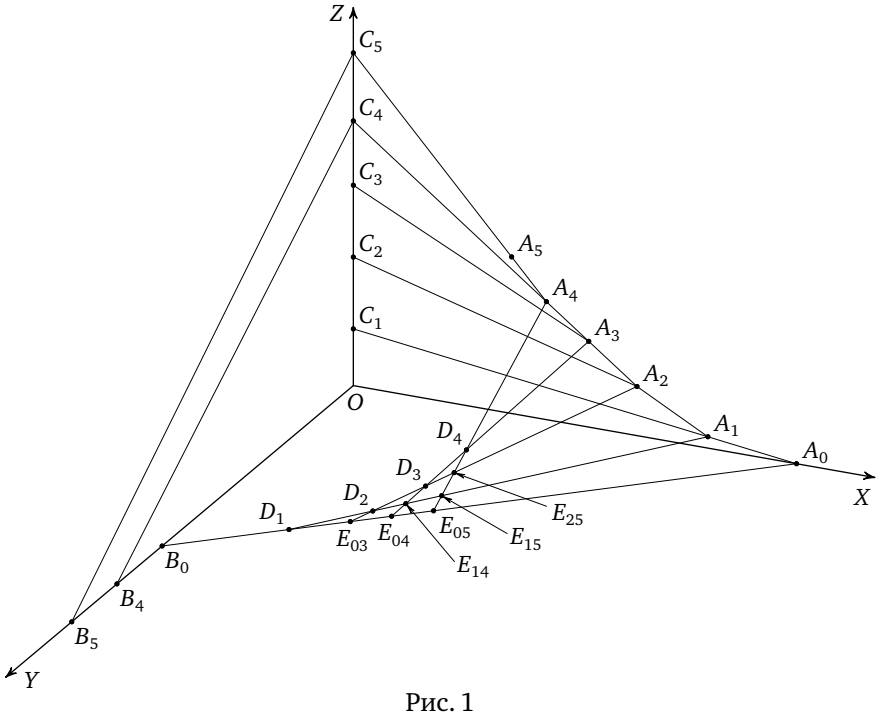


Рис. 1

требованию задачи. Обозначим точки пересечения плоскости r_k, $k > 2$, с прямыми B_0A_0, D_1A_1, D_2A_2, … через E_{0k}, E_{1k}, E_{2k}, … соответственно (см. рис. 1).

Треугольник $A_1D_1A_0$ является частью поверхности многогранника S_1. Кроме того,

$$\triangle A_1D_1A_0 \subset r_0^+ \cap r_1^+ \cap r_2^+. \tag{1}$$

Плоскости $r_3, r_4, \ldots, r_k, \ldots$ пересекают треугольник $A_1D_1A_0$ по отрезкам D_2E_{03}, $E_{14}E_{04}$, …, $E_{1k}E_{0k}$, … соответственно, причём часть треугольника слева от $E_{1k}E_{0k}$ (см. рис. 1) лежит в r_k^-, а часть справа — в r_k^+; следовательно,

$$D_2D_1E_{03} \subset S_3, \quad E_{14}D_2E_{03}E_{04} \subset r_3^+ \cap r_4^-,$$
$$E_{15}E_{14}E_{04}E_{05} \subset r_3^+ \cap r_4^+ \cap r_5^-, \quad \ldots,$$

и с учётом (1)

$$D_2D_1E_{03} \subset S_3, \quad E_{14}D_2E_{03}E_{04} \subset S_4, \quad E_{15}E_{14}E_{04}E_{05} \subset S_5, \quad \ldots$$

Поэтому S_1 имеет общую границу положительной площади с каждым из остальных S_k.

Аналогично, рассмотрев треугольники $A_2D_2A_1$, $A_3D_3A_2$, …, получаем тот же результат для многогранников $S_2, S_3, \ldots$

Таким образом, $\{S_k\}$ — бесконечная последовательность попарно смежных многогранников.

Список литературы

[1] *Besicovitch A. S.* On Crum's problem // J. London Math. Soc. 1947. Vol. 22. P. 285–287.

[2] *Erickson J.* Arbitrary large neighborly families of congruent symmetric convex 3-polytopes. http://arxiv.org/abs/math/0106095. 12.06.2001.

[3] *Rado R.* A sequence of polyhedra having intersections of specified dimensions // J. London Math. Soc. 1947. Vol. 22. P. 287–289.

[4] *Заславский А. А.* О попарно смежных многогранниках // Математическое просвещение. Сер. 3. Вып. 6. М.: МЦНМО, 2002. С. 127–129.

Абрам Самойлович Безикович (1891–1970)

Покрытие полосками

А. В. Доледенок, А. Н. Доледенок

§ 1. Введение

В настоящей статье мы обсудим некоторые избранные результаты по задаче о покрытии фигур полосками. Для простоты изложения почти все рассуждения и доказательства мы будем приводить лишь для двумерного случая. Тем не менее, многие доказательства почти дословно переносятся на случай произвольной размерности.

Полоска в $\mathbb{R}^d$ — это множество точек, заключённых между двумя параллельными гиперплоскостями (включая сами гиперплоскости). *Шириной* полоски будем называть расстояние между гиперплоскостями, её ограничивающими. Во избежание путаницы в случае $d = 3$ иногда будем называть полоску *слоем*.

Основу исследования задачи о покрытии полосками заложили работы А. Тарского (см. [17, 18]), в которых он рассматривал величину $\tau(x)$, равную наименьшему количеству частей, на которые необходимо разбить прямоугольник $x \times \frac{1}{x}$, чтобы сложить единичный квадрат. В статье [16] Х. Мёзе решил частный случай этой задачи, а именно, он показал, что $\tau(n) = n$. Тарский, используя ключевую идею из работы Мёзе, доказал следующее утверждение:

Если круг покрыт полосками, то сумма их ширин не меньше диаметра круга.

Приведём доказательство этого утверждения. Нам потребуется красивый факт, известный ещё Архимеду. Пересечём сферу и слой так, чтобы обе плоскости, ограничивающие слой, имели со сферой общие точки. Тогда площадь пересечения слоя со сферой равна $\pi\omega D$, где ω — расстояние между плоскостями, а D — диаметр сферы. Это означает, что как бы ни был расположен слой ширины ω, площадь его пересечения со сферой будет одинакова. В частности, площадь всей сферы равна πD^2.

Предположим, что круг C покрыт полосками, ширины которых равны $\omega_1, \omega_2, \ldots, \omega_n$. Рассмотрим сферу, для которой наш круг принадлежит

экваториальной плоскости. Каждой полоске поставим в соответствие такой слой, что ограничивающие его плоскости проходят через стороны полоски перпендикулярно плоскости круга (см. рис. 1). Заметим, что если полосками покрыт весь круг, то соответствующие им пересечения со сферой покроют всю сферу. Поэтому сумма площадей этих пересечений должна быть не меньше площади всей сферы, т. е.

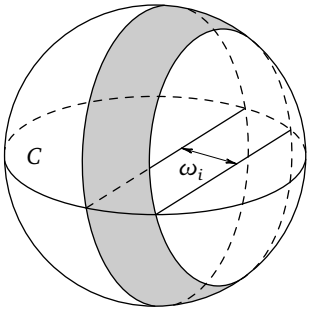


Рис. 1

$$\pi\omega_1 D + \pi\omega_2 D + \ldots + \pi\omega_n D \geqslant \pi D^2 \quad \Leftrightarrow$$
$$\Leftrightarrow \quad \omega_1 + \omega_2 + \ldots + \omega_n \geqslant D,$$

что и требовалось. Заметим также, что если внутренности каких-то двух полосок, покрывающих круг, пересекаются внутри круга, то неравенство станет строгим, т. е. суммарная ширина будет строго больше диаметра круга. Поэтому равенство суммарной ширины диаметру достигается только тогда, когда полоски лежат параллельно друг другу и стыкуются ограничивающими их прямыми, две из которых касаются круга.

Прежде чем обсуждать дальнейшие результаты, дадим несколько определений. Фигура называется *выпуклой*, если вместе с любыми двумя своими точками она целиком содержит отрезок, их соединяющий. *Шириной фигуры в направлении* $\vec{v}$ называется ширина $\omega(\vec{v})$ самой узкой полоски, перпендикулярной вектору $\vec{v}$ и содержащей данную фигуру. Иначе говоря, проведём прямую, параллельную $\vec{v}$, и спроецируем на эту прямую нашу фигуру. Длина получившегося отрезка и будет равна ширине по направлению. Например, ширина круга в любом направлении равна его диаметру, а у единичного квадрата ширина по направлению может принимать любое значение от 1 до $\sqrt{2}$.

Шириной фигуры называется наименьшая из ширин по всем направлениям, т. е. ширина самой узкой полоски, в которую можно заключить фигуру (см. рис. 2). Таким образом, ширина круга равна его диаметру, а ширина квадрата равна длине его стороны.

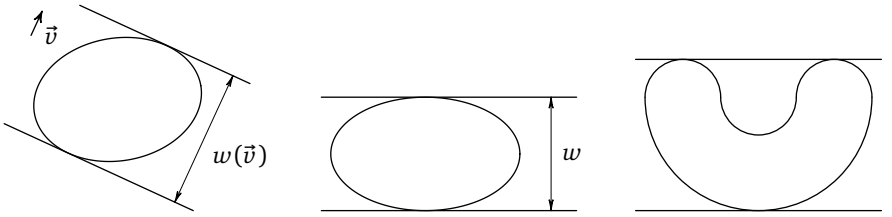


Рис. 2

В дальнейшем нам потребуется следующий факт: ширина треугольника равна длине его наименьшей высоты. Предлагаем читателю самостоятельно убедиться в его справедливости.

Вернёмся к покрытию полосками. В 1950 году Т. Банг в работе [5] доказал обобщение утверждения про покрытие круга полосками:

Теорема Банга. *Пусть выпуклое тело F в $\mathbb{R}^d$ покрыто конечным числом полосок. Тогда сумма ширин этих полосок не меньше ширины F.*

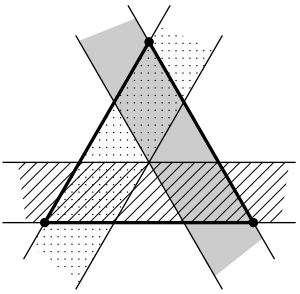


Рис. 3

Несложно убедиться, что без условия выпуклости тела теорема Банга неверна. Как и для круга, в случае произвольной фигуры имеется естественное покрытие полосками, когда полоски лежат параллельно друг другу в направлении ширины фигуры, стыкуясь ограничивающими их прямыми. Стоит отметить, что если в случае круга это был единственный способ покрыть фигуру так, чтобы суммарная ширина полосок была равна ширине фигуры, то в случае произвольной выпуклой фигуры могут существовать и другие способы. Например, правильный треугольник можно покрыть не только одной полоской ширины, равной высоте треугольника, но и тремя полосками так, как изображено на рис. 3. Ясно, что их суммарная ширина равна высоте треугольника.

Иными словами, теорема Банга говорит нам, что ничего лучше, чем класть дощечки рядом друг с другом, мы не придумаем. Мы можем придумать не хуже, но лучше — не можем.

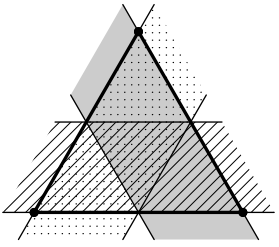


Рис. 4

Как и при доказательстве теоремы Банга для круга, в случае $d = 2$ можно было бы пытаться придумать для произвольной выпуклой фигуры такую поверхность, что любой слой фиксированной ширины высекает на ней фигуру фиксированной площади. Однако несложно показать, что для правильного треугольника такой поверхности не существует. В противном случае, если треугольник единичной ширины покрыт полосками в k слоёв (т. е. каждая точка покрыта хотя бы k полосками), их суммарная ширина должна быть не меньше k. Но на рис. 4 правильный треугольник покрыт тремя полосками, ширина каждой из которых в два раза меньше ширины треугольника, т. е. суммарная ширина меньше, чем 2. Более того, поверхность с указанным свойством существует только для круга [10]. Таким образом, для доказательства теоремы Банга в случае произвольной

фигуры необходимо использовать принципиально другие соображения. Мы приведём доказательство теоремы Банга в § 3. Доказательство теоремы Банга уже публиковалось в «Математическом просвещении» — его можно найти в статье [2].

Решив исходную задачу, Т. Банг сформулировал новую, постановку которой можно назвать более естественной, поскольку она инвариантна относительно аффинных преобразований. Нам потребуется ещё одно определение: *относительной шириной* полоски назовём отношение ширины этой полоски к ширине фигуры в направлении, перпендикулярном полоске.

Гипотеза Банга. *Если выпуклое тело F в $\mathbb{R}^d$ покрыто конечным числом полосок, то сумма относительных ширин этих полосок не меньше* 1.

Эта гипотеза является усилением теоремы Банга. Действительно, рассмотрим направление, ширина вдоль которого равна ширине фигуры. Для каждой полоски S_i из покрытия рассмотрим такую полоску S_i', что ограничивающие её гиперплоскости перпендикулярны этому направлению, а относительная ширина такая же. Понятно, что абсолютная ширина полоски S_i' не больше ширины полоски S_i. Если сумма относительных ширин хотя бы 1, то сумма ширин новых полосок не меньше ширины фигуры, что и утверждает теорема Банга. Для трёх и более полосок гипотеза Банга до сих пор не доказана. В 1991 году К. Болл доказал гипотезу Банга [4] для центрально-симметричных тел. На данный момент это является наилучшим продвижением. Подробнее речь о гипотезе Банга пойдёт в § 5.

В 2003 году В. М. Кадец доказал в работе [13] один из наиболее сильных на данный момент результатов в задаче о покрытии выпуклыми телами единичного шара. Для выпуклого тела F *вписанной сферой* назовём сферу наибольшего радиуса, которая содержится внутри F.

Теорема Кадеца. *Пусть выпуклое тело F покрыто выпуклыми телами $F_1, F_2, \ldots, F_n$. Тогда радиус вписанной сферы F не больше, чем сумма радиусов вписанных сфер $F_1, \ldots, F_n$.*

Из теоремы Кадеца вытекает теорема Банга для шара. Действительно, в качестве F нужно взять шар, а в качестве $F_1, \ldots, F_n$ — полоски. Доказательство теоремы Кадеца мы обсудим в § 4.

Теорему Банга можно интерпретировать следующим образом. Пусть в $\mathbb{R}^d$ расположено выпуклое тело и несколько полосок. Переместить полоски таким образом, чтобы покрыть ими тело, можно в том и только том случае, когда их суммарная ширина не меньше ширины тела. Вопрос становится более интригующим, если запретить поворачивать полоски, т. е. разрешить только параллельные переносы полосок. При каком условии

ими можно будет покрыть тело? Впервые эта задача была рассмотрена в работе [11].

В плоском случае для единичного круга существует такая константа c, что любая система полосок на плоскости, с суммарной шириной не меньше c, допускает покрытие круга их параллельными переносами, подробнее про это мы поговорим в § 6. В случае $d \geqslant 3$ задача не решена окончательно. Неизвестно даже, удастся ли покрыть единичный шар параллельными переносами полосок, если сумма их ширин бесконечна. Соответствующее утверждение носит название гипотезы Макаи — Паха (см. [15]).

Гипотеза Макаи — Паха. *Бесконечная последовательность полосок в пространстве $\mathbb{R}^d$ с ширинами $\omega_1, \omega_2, \ldots$ допускает покрытие пространства $\mathbb{R}^d$ параллельными переносами тогда и только тогда, когда*

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

Несложно видеть, что если параллельными переносами полосок с бесконечной суммой ширин можно покрыть единичный шар, то можно покрыть таким образом и всё пространство. Действительно, разобьём полоски на два множества таким образом, чтобы сумма ширин в каждом множестве была бесконечна. Полосками первого множества покроем единичный шар. Проделав ту же самую операцию с оставшимися полосками, их частью покроем ещё один шар радиуса 1 и т. д. Поскольку единичными шарами можно покрыть всё пространство, то и параллельными переносами полос можно покрыть всё пространство.

Необходимость расходимости ряда в гипотезе Макаи — Паха напрямую следует из теоремы Банга: если $\sum_{i=1}^{\infty} \omega_i < D$, то полосками нельзя покрыть даже шар диаметра D. Достаточность доказана только для случая $d = 2$ в работе [15]. Для $d \geqslant 3$ наилучший результат принадлежит А. Б. Купавскому и Я. Паху [14]. Из него, в частности, следует, что если ширины полос образуют гармонический ряд, то их параллельными сдвигами можно покрыть всё $\mathbb{R}^d$. Подробнее этот результат обсуждается в § 6.

§ 2. Вспомогательные утверждения

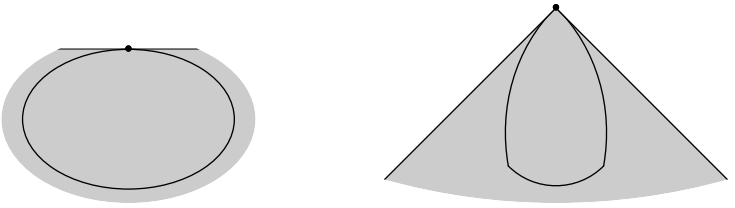
Основным объектом статьи будут выпуклые фигуры на плоскости. Нам понадобятся следующие свойства выпуклых фигур.

- *Пересечение выпуклых фигур является выпуклой фигурой.*

Действительно, если взять точки A и B внутри пересечения двух выпуклых фигур, то отрезок AB лежит и в первой фигуре, и во второй.

- *В каждой точке границы выпуклой фигуры F можно провести опорную прямую.*

Напомним, что прямая называется *опорной*, если она имеет общие точки с фигурой и вся фигура лежит по одну сторону от прямой. Возьмём произвольную точку A границы. Проведём из неё всевозможные лучи, проходящие через отличные от A точки фигуры. В силу выпуклости эти лучи заполнят полуплоскость или угол, меньший 180° (см. рис. 5). В первом случае опорной прямой будет ограничивающая полуплоскость прямая. Во втором случае подойдёт любая прямая, не проходящая внутри угла.

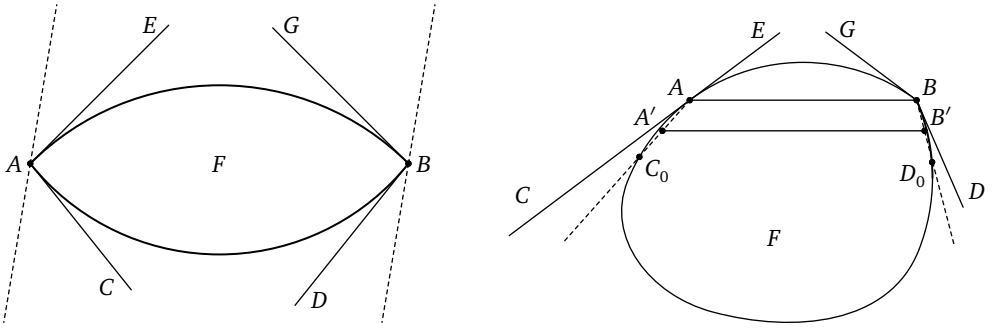


а) б)

Рис. 5

- *Через концы наибольшей из хорд выпуклой фигуры F, имеющих направление $\vec{v}$, можно провести пару параллельных опорных прямых.*

Пусть AB — наибольшая из хорд, имеющих направление $\vec{v}$ (если их несколько, то рассмотрим любую из них). Аналогично рис. 5 б) построим углы CAE и DBG, в которых содержится F (см. рис. 6 а). Докажем, что $\angle CAB + \angle DBA \leqslant 180^\circ$. Предположим противное (см. рис. 6 б). Выберем на границе F такие точки C_0 и D_0, что $\angle C_0AB + \angle D_0BA > 180^\circ$. В силу выпуклости, отрезки AC_0 и BD_0 целиком лежат внутри фигуры. Проведём прямую, параллельную AB и пересекающую отрезки AC_0 и BD_0 в точ-



а) б)

Рис. 6

ках A' и B' соответственно. Тогда длина отрезка $A'B'$ больше длины отрезка AB, что противоречит выбору AB. Следовательно, $\angle CAB + \angle DBA \leqslant 180^\circ$. Аналогично $\angle EAB + \angle GBA \leqslant 180^\circ$. Теперь существование нужных опорных прямых очевидно.

§ 3. Доказательство теоремы Банга

Мы приведём доказательство теоремы Банга для плоскости от противного. В случае произвольной размерности доказательство полностью аналогично. Предположим, что фигура F покрыта полосками $S_1, S_2, \ldots, S_n$, сумма ширин которых меньше ширины фигуры F. Полоске S_i сопоставим вектор $\vec{v}_i$, перпендикулярный ограничивающим её прямым (из двух возможных направлений вектора выберем произвольное) и по длине равный половине ширины S_i. Рассмотрим множество из 2^n точек:

$$O + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n),$$

где $\lambda > 1$ таково, что сумма ширин полосок $S_1, \ldots, S_n$, умноженная на λ, меньше ширины фигуры F.

Доказательство теоремы Банга будет состоять из двух частей. На первом шаге мы докажем, что можно выбрать точку O таким образом, что все 2^n точек будут лежать внутри фигуры F. На втором шаге мы докажем, что все эти точки не могут быть покрыты полосками. Полученное противоречие завершит доказательство.

Лемма 1. *Пусть F — выпуклая фигура, ω — её ширина, F' — фигура, полученная из F параллельным переносом на вектор $\vec{v}$. Тогда ширина пересечения $F \cap F'$ не меньше, чем $\omega - |\vec{v}|$.*

Доказательство. Пусть $|\vec{v}| < \omega$, в противном случае утверждение очевидно. Пусть AB — наибольшая из хорд фигуры F, имеющих направление $\vec{v}$ (см. рис. 7). Проведём через точки A и B параллельные опорные

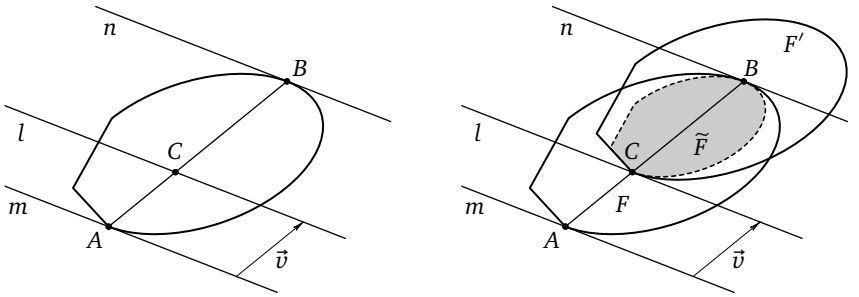


Рис. 7

прямые m и n. Пусть прямая l получается из прямой m параллельным переносом на вектор $\vec{v}$, C — точка пересечения прямых l и AB. Без ограничения общности, l пересекается с фигурой F.

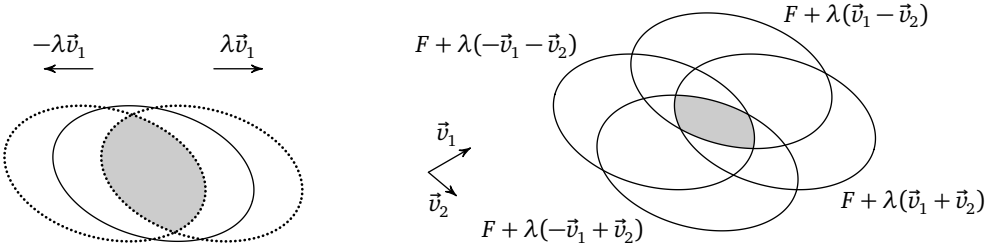
Рассмотрим фигуру $\widetilde{F}$, полученную из F гомотетией с центром в точке B и коэффициентом $\frac{BC}{BA}$. Очевидно, что $\widetilde{F}$ лежит внутри F. Заметим, что $\widetilde{F}$ можно получить из F' гомотетией с центром в точке C и таким же коэффициентом, поэтому $\widetilde{F}$ лежит и внутри F'. Получаем, что ширина пересечения не меньше чем $\frac{BC}{BA} \cdot \omega$. Воспользовавшись тем, что $|\vec{v}| = AC$, получим цепочку равносильных неравенств:

$$\frac{BC}{BA} \cdot \omega \geqslant \omega - |\vec{v}| \quad \Leftrightarrow \quad |\vec{v}| \geqslant \omega \cdot \frac{AC}{AB} \quad \Leftrightarrow \quad AC \geqslant \omega \cdot \frac{AC}{AB} \quad \Leftrightarrow \quad AB \geqslant \omega.$$

Последнее неравенство верно, поскольку ω — ширина F. □

Лемма 2. *Рассмотрим 2^n сдвигов фигуры F: $F + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$. Пересечение этих фигур не пусто, причём ширина общего пересечения не меньше, чем $\omega - 2\lambda(|\vec{v}_1| + |\vec{v}_2| + \ldots + |\vec{v}_n|)$.*

Доказательство. Докажем утверждение индукцией по числу векторов n. База при $n = 1$ вытекает из леммы 1 (фигура $F + \lambda\vec{v}_1$ получается из $F - \lambda\vec{v}_1$ переносом на вектор $2\lambda\vec{v}_1$, см. рис. 8).



а) б)

Рис. 8

Докажем переход. Обозначим через F_{k-1} пересечение 2^{k-1} фигур

$$F + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_{k-1}).$$

По предположению индукции ширина F_{k-1} не меньше чем

$$\omega - 2\lambda(|\vec{v}_1| + |\vec{v}_2| + \ldots + |\vec{v}_{k-1}|).$$

Пусть F_k — пересечение фигур $F_{k-1} + \lambda\vec{v}_k$ и $F_{k-1} - \lambda\vec{v}_k$. Покажем, что F_k лежит в пересечении всех 2^k фигур (отмечено серым на рис. 8). Поскольку ширина F_k не меньше, чем ширина F_{k-1}, уменьшенная на $2\lambda|\vec{v}_k|$, то тем самым утверждение леммы будет доказано.

Рассмотрим фигуру $F + \lambda(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_k\vec{v}_k)$ для некоторых $\alpha_1, \alpha_2, \ldots, \alpha_k \in \{-1, 1\}$. Тогда

$$F_{k-1} \subset F + \lambda(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_{k-1}\vec{v}_{k-1}),$$

откуда

$$F_{k-1} + \lambda\alpha_k\vec{v}_k \subset F + \lambda(\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_{k-1}\vec{v}_{k-1} + \alpha_k\vec{v}_k).$$

Но $F_k \subset F_{k-1} + \lambda\alpha_k\vec{v}_k$, откуда получаем требуемое. □

Рассмотрим некоторую точку O, лежащую в пересечении 2^n фигур из леммы 2. Выберем произвольные $\alpha_1, \alpha_2, \ldots, \alpha_n \in \{-1, 1\}$. Согласно выбору точки O, она принадлежит фигуре $F + \lambda(-\alpha_1\vec{v}_1 - \alpha_2\vec{v}_2 - \ldots - \alpha_n\vec{v}_n)$. Сделав параллельный перенос на вектор $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_n\vec{v}_n$, получим, что

$$O + (\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \ldots + \alpha_n\vec{v}_n) \in F.$$

Поскольку изначально мы взяли произвольные $\alpha_1, \ldots, \alpha_n \in \{-1, 1\}$, то выполнен первый шаг, а именно, мы нашли такую точку O, что 2^n точек $O + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$ принадлежат фигуре F.

Перейдём ко второму шагу доказательства. Достаточно доказать следующую лемму.

Лемма Банга. *Для каждой полоски S_i из множества полосок $S_1, S_2, \ldots, S_n$ зафиксируем вектор $\vec{v}_i$, перпендикулярный прямым, ограничивающим эту полоску, и по длине равный половине ширины S_i, а также зафиксируем произвольное число $\lambda > 1$. Тогда для любой точки O хотя бы одна из 2^n точек множества $O + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$ не покрыта полосками.*

Мы приведём два доказательства леммы Банга. Первое доказательство основано на работе [7].

Первое доказательство леммы Банга. Рассмотрим плоскость γ, в которой находятся полоски, и произвольную точку O в ней. Пусть $\vec{e}$ — единичный вектор нормали к γ. Рассмотрим такую точку Y, что $\overrightarrow{OY} = t\vec{e}$ для некоторого числа $t > 0$. Каждой полоске S_i сопоставим слой S_i' следующим образом. Плоскость, проходящая через точку Y и срединную прямую полоски S_i, является срединной плоскостью слоя S_i', а плоскости, ограничивающие слой, проходят через прямые, ограничивающие S_i. Обозначим через $\vec{v}_i{}'$ перпендикулярный S_i' вектор, длина которого равна половине ширины S_i', причём $(\vec{v}_i, \vec{v}_i{}') > 0$ (см. рис. 9).

Докажем, что слои $S_1', S_2', \ldots, S_n'$ не покрывают хотя бы одну из 2^n точек $O + \lambda(\pm\vec{v}_1{}' \pm \vec{v}_2{}' \pm \ldots \pm \vec{v}_n{}')$. Выберем среди них наиболее удалённую от Y

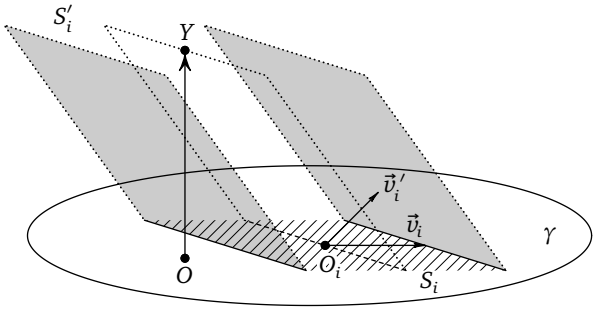


Рис. 9

точку X. Предположим, что она покрыта слоем S_i'. Заменим в разложении вектора $\overrightarrow{OX}$ знак перед вектором $\vec{v}_i{}'$ на противоположный, полученную точку обозначим через X_1. Отрезок XX_1 перпендикулярен срединной плоскости слоя S_i', а его длина равна $2\lambda|\vec{v}_i{}'|$, поэтому точка X_1 не лежит в слое S_i', а следовательно, расположена от Y дальше, чем X. Полученное противоречие показывает, что наиболее удалённая от Y точка не покрыта слоями. Более того, из доказательства несложно видеть, что все точки шара с центром в точке X и радиусом $(\lambda-1)d'$, где $d'=\min(|\vec{v}_1{}'|, |\vec{v}_2{}'|, \ldots, |\vec{v}_n{}'|)$, также не покрыты слоями.

Устремим t к бесконечности. Очевидно, что при этом $\vec{v}_i{}' \to \vec{v}_i$ для всех i, поэтому все точки множества $O + \lambda(\pm\vec{v}_1{}' \pm \vec{v}_2{}' \pm \ldots \pm \vec{v}_n{}')$ стремятся к соответствующим точкам множества $O + \lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$ и $d' \to d$, где $d = \min(|\vec{v}_1|, |\vec{v}_2|, \ldots, |\vec{v}_n|)$. Поэтому при достаточно большом t каждый из 2^n шаров с центрами в точках множества $O + \lambda(\pm\vec{v}_1{}' \pm \vec{v}_2{}' \pm \ldots \pm \vec{v}_n{}')$ и радиусами $(\lambda-1)d'$ содержит соответствующую центру точку из множества $O+\lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$. Поскольку мы показали, что хотя бы один из этих шаров не покрыт слоями $S_1', \ldots, S_n'$, то и одна из точек множества $O+\lambda(\pm\vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n)$ не покрыта слоями, а следовательно, и полосками $S_1, \ldots, S_n$. Лемма, а вместе с ней и теорема Банга, доказана. □

Приведём второе доказательство леммы Банга. Но предварительно проделаем выкладки в обозначениях первого доказательства. Пусть

$$\overrightarrow{OX} = \lambda \sum_{i=1}^{n} \alpha_i \vec{v}_i{}',$$

где $\alpha_i \in \{-1, 1\}$ для любого i. Распишем квадрат длины отрезка XY:

$$|XY|^2 = \left(\lambda \sum_{i=1}^{n} \alpha_i \vec{v}_i{}' - t\vec{e}\right)^2 = \lambda^2 \sum_{i,j=1}^{n} (\alpha_i \vec{v}_i{}', \alpha_j \vec{v}_j{}') - 2\lambda \sum_{i=1}^{n} (\alpha_i \vec{v}_i{}', t\vec{e}) + t^2.$$

Обозначим через O_i проекцию точки O на срединную прямую полоски S_i (см. рис. 10). Из геометрических соображений легко получить, что

$$(\vec{v}_i{}', t\vec{e}) = \pm\frac{t^2\cdot|\vec{v}_i|\cdot|OO_i|}{t^2+|OO_i|^2},$$

где перед дробью стоит знак «+», если векторы $\overrightarrow{OO_i}$ и $\vec{v}_i$ сонаправлены, и знак «—» в противном случае. Отсюда

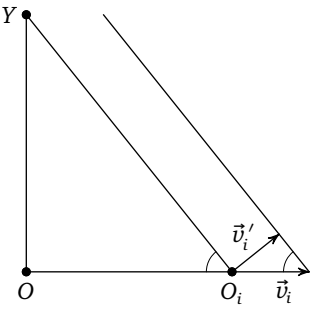


Рис. 10

$$(\alpha_i\vec{v}_i{}', t\vec{e}) = \pm\alpha_i\frac{t^2\cdot|\vec{v}_i|\cdot|OO_i|}{t^2+|OO_i|^2} = (\overrightarrow{OO_i}, \alpha_i\vec{v}_i)\cdot\frac{t^2}{t^2+|OO_i|^2}.$$

В итоге получаем, что

$$|XY|^2 = \lambda^2\sum_{i,j=1}^{n}(\alpha_i\vec{v}_i{}', \alpha_j\vec{v}_j{}') - 2\lambda\sum_{i=1}^{n}(\overrightarrow{OO_i}, \alpha_i\vec{v}_i)\cdot\frac{t^2}{t^2+|OO_i|^2} + t^2.$$

Заметим, что предел выражения $\frac{1}{\lambda}(|XY|^2 - |OY|^2)$ при $t\to+\infty$ будет конечным, а именно

$$\lim_{t\to+\infty}\frac{1}{\lambda}(XY^2 - OY^2) = \lim_{t\to+\infty}\frac{1}{\lambda}(XY^2 - t^2) = \lambda\sum_{i,j=1}^{n}(\alpha_i\vec{v}_i, \alpha_j\vec{v}_j) - 2\sum_{i=1}^{n}(\overrightarrow{OO_i}, \alpha_i\vec{v}_i).$$

Согласно выбору точки X, длина отрезка XY — наибольшая возможная, поэтому выражение $\frac{1}{\lambda}(XY^2 - OY^2)$ также принимает наибольшее возможное значение. Отсюда получаем

Второе доказательство леммы Банга. Определим на множестве наборов $(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)\in\{-1, 1\}^n$ функционал

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \lambda\sum_{i,j=1}^{n}(\varepsilon_i\vec{v}_i, \varepsilon_j\vec{v}_j) - 2\sum_{i=1}^{n}(\overrightarrow{OO_i}, \varepsilon_i\vec{v}_i).$$

Пусть функционал достигает максимума на наборе $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Докажем, что точка X, определяемая равенством

$$\overrightarrow{OX} = \lambda\sum_{i=1}^{n}\alpha_i\vec{v}_i,$$

не покрыта полосками. Пусть X покрыта некоторой полоской (без ограничения общности, полоской S_1). Изменим знак при векторе $\vec{v}_1$ на противоположный и докажем, что на полученной точке X_1 функционал принимает большее значение. Выделим в выражении для $f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ все

слагаемые, не зависящие от $\vec{v}_1$:

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \lambda \sum_{i,j=2}^{n} (\varepsilon_i \vec{v}_i, \varepsilon_j \vec{v}_j) + 2\lambda \sum_{i=2}^{n} (\varepsilon_1 \vec{v}_1, \varepsilon_i \vec{v}_i) + \\ + \lambda |\vec{v}_1|^2 - 2 \sum_{i=2}^{n} (\overrightarrow{OO_i}, \varepsilon_i \vec{v}_i) - 2(\overrightarrow{OO_1}, \varepsilon_1 \vec{v}_1).$$

Тогда

$$f(\alpha_1, \alpha_2, \ldots, \alpha_n) - f(-\alpha_1, \alpha_2, \ldots, \alpha_n) = \\ = \left(2\lambda \sum_{i=2}^{n} (\alpha_1 \vec{v}_1, \alpha_i \vec{v}_i) - 2\lambda \sum_{i=2}^{n} (-\alpha_1 \vec{v}_1, \alpha_i \vec{v}_i) \right) - \\ - \left(2(\overrightarrow{OO_1}, \alpha_1 \vec{v}_1) - 2(\overrightarrow{OO_1}, -\alpha_1 \vec{v}_1) \right) = 4\lambda \sum_{i=2}^{n} (\alpha_1 \vec{v}_1, \alpha_i \vec{v}_i) - 4(\overrightarrow{OO_1}, \alpha_1 \vec{v}_1) = \\ = 4\left(\alpha_1 \vec{v}_1, \lambda \sum_{i=2}^{n} \alpha_i \vec{v}_i - \overrightarrow{OO_1} \right) = 4\left(\alpha_1 \vec{v}_1, \lambda \sum_{i=1}^{n} \alpha_i \vec{v}_i - \overrightarrow{OO_1} \right) - 4\lambda |\vec{v}_1|^2.$$

Поскольку $\lambda \sum_{i=1}^{n} \alpha_i \vec{v}_i - \overrightarrow{OO_1}$ — вектор с началом на срединной прямой полоски S_1 и концом внутри S_1 (см. рис. 11), имеем

$$\left(\alpha_1 \vec{v}_1, \lambda \sum_{i=1}^{n} \alpha_i \vec{v}_i - \overrightarrow{OO_1} \right) < \lambda |\vec{v}_1|^2 \quad \Rightarrow$$

$$\Rightarrow \quad f(\alpha_1, \alpha_2, \ldots, \alpha_n) < f(-\alpha_1, \alpha_2, \ldots, \alpha_n),$$

противоречие. Таким образом, лемма Банга доказана. □

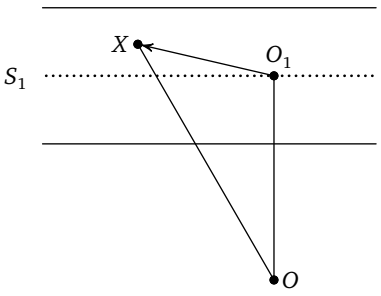


Рис. 11

§ 4. Теорема Кадеца

В этом параграфе мы докажем теорему Кадеца для плоского случая. В случае размерности пространства $d > 2$ доказательство аналогично.

Теорема Кадеца. *Пусть выпуклое тело F покрыто выпуклыми телами $F_1, F_2, \ldots, F_n$. Тогда радиус вписанной сферы F не больше, чем сумма радиусов вписанных сфер $F_1, \ldots, F_n$.*

Нам понадобятся следующие свойства вписанной окружности.

1. *У всякой выпуклой фигуры либо есть одна вписанная окружность, либо их бесконечно много.*

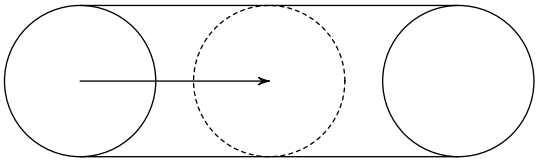
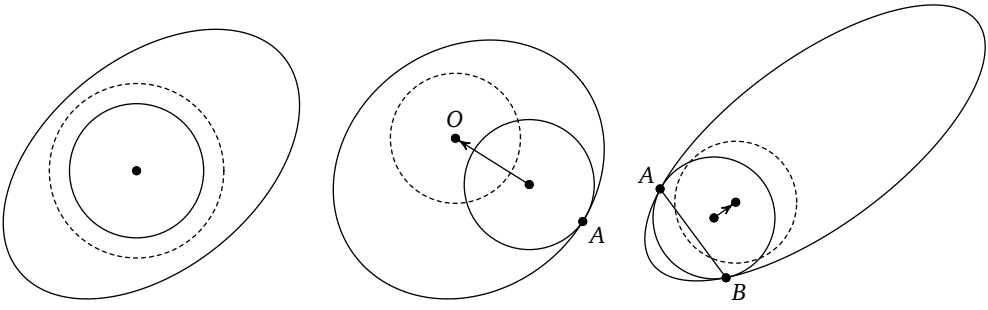


Рис. 12

Пусть есть две таких окружности. Проведём отрезки их общих внешних касательных, заключённые между точками касания (см. рис. 12). Фигура, полученная объединением окружностей и этих отрезков, лежит целиком внутри F в силу выпуклости. Поэтому если параллельно сдвинуть одну окружность в направлении другой, то образ также будет лежать внутри F. Итого получаем, что вписанных окружностей бесконечно много, причём все они касаются параллельных прямолинейных кусков границы F.

2. *Среди точек касания вписанной окружности с границей фигуры есть либо две диаметрально противоположных точки, либо три точки, которые являются вершинами остроугольного треугольника.*

Если у окружности нет общих точек с границей, то, немного увеличив её радиус (не меняя центра), получим окружность большего радиуса, также лежащую внутри нашей фигуры (см. рис. 13 а). Если у окружности одна общая точка A с границей, то сдвинем окружность параллельно прямой OA (где O — центр окружности) так, чтобы у новой окружности общих точек с границей не было (см. рис. 13 б). Полученную окружность также можно увеличить. Пусть окружность имеет две или больше общих точек с границей, причём существует дуга AB, большая 180° и не содержащая других граничных точек. Тогда сдвинем нашу окружность в сторону дуги по перпендикуляру к хорде AB, после чего её опять можно увеличить (см. рис. 13 в).



а) б) в)

Рис. 13

3. *Рассмотрим граничную точку, в которой вписанная окружность касается фигуры F. Тогда в этой точке касательная к окружности является опорной прямой для F.*

Рассмотрим опорную прямую в точке касания. Она является опорной и для окружности, однако единственная опорная прямая к окружности — это касательная.

Доказательство теоремы Кадеца. Обозначим через r радиус окружности, вписанной в фигуру F, через O — центр этой окружности. Для всех i обозначим через r_i радиус окружности, вписанной в фигуру F_i, а через O_i — её центр (если F_i — это полоска, то выберем любую из вписанных окружностей). Предположим, что условие теоремы неверно и $\sum_{i=1}^{n} r_i < r$. Вписанный в F круг также покрыт фигурами $F_1, \ldots, F_n$, поэтому можно считать, что F — это круг. Также можно считать, что каждая фигура F_i — это треугольник или полоска. Действительно, по свойству 2 существуют либо две диаметрально противоположные точки касания вписанной окружности с границей F_i, либо три точки касания, образующие остроугольный треугольник. Проведём касательные в этих точках, они образуют полоску или треугольник. При этом образовавшаяся фигура содержит F_i и имеет такой же радиус вписанной окружности.

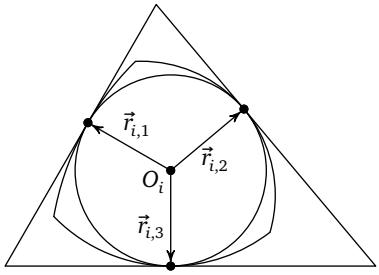


Рис. 14

Пусть $\vec{r}_{i,1}, \vec{r}_{i,2}, \vec{r}_{i,3}$ — векторы, проведённые из точки O_i в точки касания окружности с границей (см. рис. 14). Если точек касания не три, а две, то векторов также будет два.

Рассмотрим случай, когда каждая фигура F_i является треугольником (случай, когда в покрытии присутствуют полоски, доказывается аналогично). Рассмотрим 3^n точек вида

$$O + \lambda \sum_{i=1}^{n} \vec{r}_{i,\varepsilon_i},$$

где $\varepsilon_i \in \{1, 2, 3\}$, а $\lambda > 1$ таково, что $\lambda \sum_{i=1}^{n} r_i < r$. В силу выбора λ все полученные точки будут лежать внутри F.

Как и во втором доказательстве леммы Банга, рассмотрим функционал

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \lambda \sum_{i,j=1}^{n} (\vec{r}_{i,\varepsilon_i}, \vec{r}_{j,\varepsilon_j}) - 2 \sum_{i=1}^{n} (\overrightarrow{OO_i}, \vec{r}_{i,\varepsilon_i}),$$

где $\varepsilon_i \in \{1, 2, 3\}$ для всех i. Пусть функционал достигает максимума на наборе $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Докажем, что точка X_1, определяемая равенством

$$\overrightarrow{OX_1} = \lambda \sum_{i=1}^{n} \vec{r}_{i,\alpha_i},$$

не покрыта фигурами $F_1, \ldots, F_n$.

Предположим противное: пусть точка X_1 покрыта каким-нибудь треугольником (без ограничения общности, треугольником F_1). Для определённости будем считать, что $\alpha_1 = 1$. Изменим в разложении вектора $\overrightarrow{OX_1}$ первое слагаемое, а именно, рассмотрим точки X, X_2, X_3, определённые равенствами:

$$\overrightarrow{OX} = \lambda \sum_{i=2}^{n} \vec{r}_{i,\alpha_i}, \quad \overrightarrow{OX_2} = \lambda\vec{r}_{1,2} + \lambda \sum_{i=2}^{n} \vec{r}_{i,\alpha_i}, \quad \overrightarrow{OX_3} = \lambda\vec{r}_{1,3} + \lambda \sum_{i=2}^{n} \vec{r}_{i,\alpha_i}.$$

Выделим в $f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ слагаемые, не зависящие от $\vec{r}_{1,\varepsilon_1}$:

$$f(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \lambda \sum_{i,j=2}^{n} (\vec{r}_{i,\varepsilon_i}, \vec{r}_{j,\varepsilon_j}) + 2\lambda \sum_{i=2}^{n} (\vec{r}_{1,\varepsilon_1}, \vec{r}_{i,\varepsilon_i}) + \\ + \lambda|\vec{r}_{1,\varepsilon_1}|^2 - 2\sum_{i=2}^{n} (\overrightarrow{OO_i}, \vec{r}_{i,\varepsilon_i}) - 2(\overrightarrow{OO_1}, \vec{r}_{1,\varepsilon_1}).$$

Запишем разность значений функционала для точек X_1 и X_2:

$$\begin{aligned} &f(1, \alpha_2, \ldots, \alpha_n) - f(2, \alpha_2, \ldots, \alpha_n) = \\ &\quad = 2\lambda\Big(\sum_{i=2}^{n} (\vec{r}_{1,1}, \vec{r}_{i,\alpha_i}) - \sum_{i=2}^{n} (\vec{r}_{1,2}, \vec{r}_{i,\alpha_i}) \Big) - 2\big((\overrightarrow{OO_1}, \vec{r}_{1,1}) - (\overrightarrow{OO_1}, \vec{r}_{1,2}) \big) = \\ &\quad = 2\lambda\Big(\vec{r}_{1,1} - \vec{r}_{1,2},\ \sum_{i=2}^{n} \vec{r}_{i,\alpha_i} \Big) - 2(\vec{r}_{1,1} - \vec{r}_{1,2}, \overrightarrow{OO_1}) = \\ &\quad = 2\Big(\vec{r}_{1,1} - \vec{r}_{1,2},\ \lambda \sum_{i=2}^{n} \vec{r}_{i,\alpha_i} - \overrightarrow{OO_1} \Big). \end{aligned}$$

Рассмотрим треугольник F_1, в котором лежит точка X_1, обозначим его вершины через A, B, C (см. рис. 15). Пусть точки касания его вписанной окружности с центром O_1 со сторонами AB, AC и BC — это точки C_1, B_1 и A_1 соответственно, причём векторы $\overrightarrow{O_1A_1}$, $\overrightarrow{O_1B_1}$ и $\overrightarrow{O_1C_1}$ сонаправлены векторам $\vec{r}_{1,1}$, $\vec{r}_{1,2}$ и $\vec{r}_{1,3}$ соответственно.

Тогда

$$2\Big(\vec{r}_{1,1} - \vec{r}_{1,2},\ \lambda \sum_{i=2}^{n} \vec{r}_{i,\alpha_i} - \overrightarrow{OO_1} \Big) = 2(\overrightarrow{B_1A_1}, \overrightarrow{O_1X}).$$

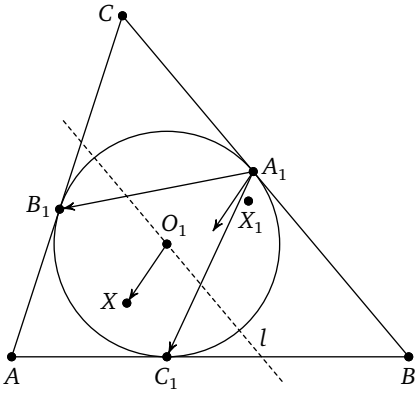


Рис. 15

Аналогично получается равенство

$$f(1, \alpha_2, \ldots, \alpha_n) - f(3, \alpha_2, \ldots, \alpha_n) = 2(\overrightarrow{C_1A_1}, \overrightarrow{O_1X}).$$

Поскольку значение функционала наибольшее в точке X_1, обе разности неотрицательны, т. е. углы между вектором $\overrightarrow{O_1X}$ и векторами $\overrightarrow{A_1B_1}$ и $\overrightarrow{A_1C_1}$ неострые. Проведём через точку O_1 прямую l параллельно стороне BC. Поскольку точка X_1 лежит внутри треугольника, точка X лежит в той же полуплоскости относительно прямой l, что и точка A. Таким образом, если отложить вектор $\overrightarrow{O_1X}$ от точки A_1, то он будет направлен в ту же полуплоскость относительно прямой BC, что и векторы $\overrightarrow{A_1B_1}$ и $\overrightarrow{A_1C_1}$. Однако векторы $\overrightarrow{A_1B_1}$ и $\overrightarrow{A_1C_1}$ разбивают развёрнутый угол BA_1C на три острых угла (что несложно получить, выразив эти углы через углы треугольника), т. е. вектор $\overrightarrow{O_1X}$ образует хотя бы с одним из них острый угол. Противоречие. □

Приведённое доказательство теоремы Кадеца идейно полностью повторяет второе доказательство леммы Банга. Заметим, что первое доказательство леммы Банга с поднятием и рассмотрением наиболее удалённой точки также можно модифицировать для доказательства теоремы Кадеца.

В плоском случае теорема также доказана в работе [6]. Доказательство идейно похоже на доказательство теоремы Банга для круга. Оно опирается на следующее утверждение.

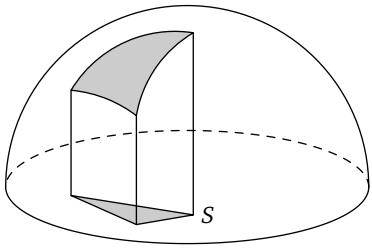


Рис. 16

Пусть B — круг на плоскости, S — пересекающая его полоска или треугольник. Обозначим через r радиус окружности, вписан-

ной в фигуру $B \cap S$. Рассмотрим такую полусферу, что B является её экваториальным сечением. Спроецируем фигуру $B \cap S$ на полусферу, проведя через каждую точку фигуры прямую, перпендикулярную плоскости круга, как показано на рис. 16. *Тогда площадь полученной проекции не превосходит $2\pi r$, причём равенство достигается, только если S — полоса, проходящая через B.*

§ 5. Гипотезы Банга и Дэвенпорта

В этом разделе речь пойдёт о гипотезе Банга.

Гипотеза Банга. *Пусть выпуклое тело F в $\mathbb{R}^d$ покрыто конечным числом полосок. Тогда сумма относительных ширин этих полосок не меньше* 1.

Докажем гипотезу Банга в плоском случае для двух полосок (для одной полоски утверждение очевидно). Пусть фигуру F покрывают две полоски. Если они параллельны, то утверждение очевидно. Иначе полоски в пересечении образуют параллелограмм, обозначим его $PQRT$. Вершины этого параллелограмма являются внешними или граничными точками F. Действительно, если P — внутренняя точка, угол при P, на рис. 17 отмеченный дугой, содержит точки фигуры, но не покрыт полосками.

Через вершины параллелограмма можно провести прямые, не пересекающие фигуру F. Действительно, если P — граничная точка F, то подойдёт опорная прямая. Если же P — внешняя точка для F, то рассмотрим ближайшую к P точку фигуры. Проведём через P прямую, параллельную опорной в этой точке. Эта прямая будет искомой.

Проведём в каждой вершине параллелограмма прямую, не пересекающую нашу фигуру. Эти прямые образуют четырёхугольник $ABCD$, который

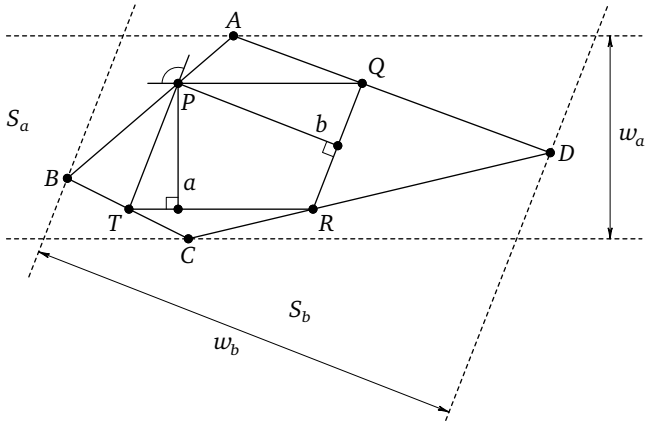


Рис. 17

целиком содержит F (см. рис. 17), т. е. его ширина по каждому направлению не меньше ширины F по этому направлению.

Пусть ширины полосок S_a и S_b равны a и b, ширина четырёхугольника в соответствующих направлениях равна ω_a и ω_b соответственно. Обозначим $\alpha = a/\omega_a$, $\beta = b/\omega_b$, площадь параллелограмма $PQRT$ обозначим через S. Рассмотрим два треугольника APQ и CTR, которые покрывает полоска S_b, но не покрывает полоска S_a. Расписав площади этих треугольников и параллелограмма через высоты и основания, получим, что отношение суммарной площади этих треугольников к площади параллелограмма равно

$$\frac{\omega_a - a}{2a} = \frac{1}{2\alpha} - \frac{1}{2}.$$

Аналогично отношение суммарной площади треугольников BPT и DQR к площади параллелограмма равно $\frac{1}{2\beta} - \frac{1}{2}$. Тогда отношение площади четырёхугольника $ABCD$ к площади параллелограмма $PQRT$ равно

$$\left(\frac{1}{2\alpha} - \frac{1}{2}\right) + \left(\frac{1}{2\beta} - \frac{1}{2}\right) + 1 = \frac{1}{2\alpha} + \frac{1}{2\beta}.$$

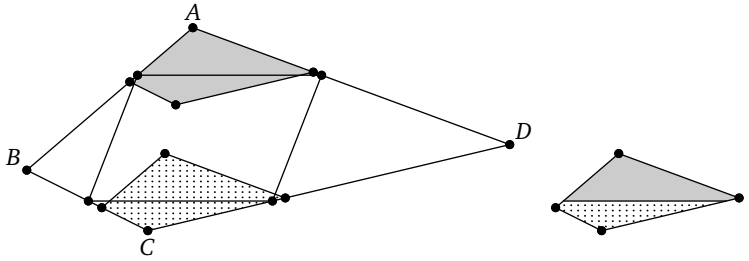


Рис. 18

Рассмотрим образы четырёхугольника $ABCD$ при гомотетиях относительно точек A и C с коэффициентом $1-\alpha$. Нетрудно понять, что части этих образов, лежащие внутри треугольников APQ и CRT, при совмещении дадут в точности четырёхугольник, гомотетичный $ABCD$ с коэффициентом $1-\alpha$ (см. рис. 18). Отсюда следует, что суммарная площадь этих треугольников не меньше площади $ABCD$, умноженной на $(1-\alpha)^2$. Получаем неравенство

$$\frac{1}{2\alpha} - \frac{1}{2} \geqslant (1-\alpha)^2\left(\frac{1}{2\alpha} + \frac{1}{2\beta}\right) \quad \Leftrightarrow \quad \frac{1-\alpha}{\alpha} \geqslant (1-\alpha)^2\left(\frac{1}{\alpha} + \frac{1}{\beta}\right) \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \frac{\beta}{\alpha} \geqslant (1-\alpha)\left(1 + \frac{\beta}{\alpha}\right) \quad \Leftrightarrow \quad \alpha + \beta \geqslant 1,$$

тем самым гипотеза Банга для двух полосок доказана.

В общем случае гипотеза Банга не доказана. Известно, что она равносильна следующей гипотезе, возникшей из задач диофантовых приближений (см. [8]).

Гипотеза Дэвенпорта. *Выпуклое тело F разрезано n гиперплоскостями на части. Тело F' получено из тела F гомотетией с коэффициентом $1/(n+1)$. Тогда существует такой параллельный перенос тела F', что он содержится в теле F и не пересекает ни одной из этих гиперплоскостей.*

Докажем, что для единичного круга B гипотезы Банга и Дэвенпорта равносильны (доказательство общего утверждения о равносильности использует ту же идею, см. [3]). Для единичного круга гипотеза Банга равносильна теореме Банга, поскольку ширина круга в каждом направлении одинакова.

Пусть верна гипотеза Банга. Рассмотрим концентрический с B круг B' радиуса $n/(n+1)$. Для каждой прямой l, разрезающей B, рассмотрим такую полоску ширины $2/(n+1)$, что ограничивающие её прямые симметричны относительно l (см. рис. 19). Заметим, что если найдётся точка в B', не попавшая внутрь ни одной из построенных полосок, то круг радиуса $1/(n+1)$ с центром в этой точке будет лежать внутри исходного круга B и не будет пересекать секущие прямые (но, возможно, будет касаться). Суммарная ширина полосок равна $2n/(n+1)$, т. е. диаметру B'. Поэтому если полоски не параллельны друг другу, то по теореме Банга найдётся непокрытая точка. Если же полоски параллельны друг другу, то гипотеза Дэвенпорта, очевидно, верна: если единичную окружность пересечь n параллельными прямыми, то ширина хотя бы одной части будет не меньше $2/(n+1)$, т. е. в неё можно будет вписать круг радиуса $1/(n+1)$.

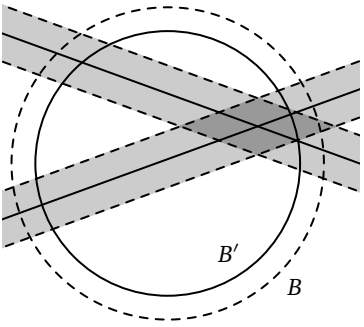


Рис. 19

Обратно, пусть верна гипотеза Дэвенпорта. Предположим, что гипотеза Банга неверна, т. е. нам удалось покрыть единичный круг полосками, сумма относительных ширин которых меньше 1 (соответственно, сумма абсолютных ширин меньше 2). Увеличим ширину каждой полоски так, чтобы она стала рациональной, но при этом сумма ширин по-прежнему была бы меньше 2 и выражалась несократимой дробью, числитель которой хотя бы на 2 больше знаменателя. Пусть ширины полосок равны $\frac{p_1}{q_1}, \frac{p_2}{q_2}, \ldots, \frac{p_n}{q_n}$. Теперь разобьём каждую полоску на меньшие полоски ширины $1/N$ каждая, где $N = q_1 q_2 \ldots q_n$. Сложим дроби (получим суммарную

ширину получившихся полосок) и приведём их к общему знаменателю:

$$\frac{p_1}{q_1}+\frac{p_2}{q_2}+\ldots+\frac{p_n}{q_n}=\frac{p_1q_2\ldots q_n+q_1p_2q_3\ldots q_n+\ldots+q_1\ldots q_{n-1}p_n}{q_1q_2\ldots q_n}=\frac{k}{N}.$$

Числитель этой дроби равен количеству полосок нового разбиения. Выбор ширин гарантирует нам, что $k<N-1$. Рассмотрим ось симметрии каждой из полосок, для этих прямых применим гипотезу Дэвенпорта. Получим, что существует круг радиуса $1/(k+1)$, целиком лежащий внутри B и не пересекающий ни одну из прямых. Поскольку $k+1<N$, центр круга удалён от каждой прямой на расстояние, большее чем $1/N$. Получаем, что он не покрыт полосками, противоречие.

§ 6. Гипотеза Макаи — Паха

Перейдём к задаче о покрытии тел параллельными переносами полосок. Дано выпуклое тело и некоторый набор полосок. Требуется выяснить, при каком условии на ширины полосок можно гарантированно покрыть тело параллельными переносами этих полосок. Принципиальное отличие этой задачи от задачи, изученной Т. Бангом, в том, что вращать полоски запрещено. Задача впервые обсуждалась в работе [11]. Мы ограничимся рассмотрением случаев размерности $d=2$ и $d=3$.

Вначале докажем, что для некоторой константы c единичный круг можно покрыть параллельными переносами любой системы полосок на плоскости суммарной ширины не меньше c. Эта задача была предложена М. Смуровым в 1997 году на Московской математической олимпиаде в качестве последней задачи в варианте 11 класса. По мотивам этой задачи в журнале «Квант» была опубликована статья [1]. Формулировка задачи:

На плоскости дано конечное число полос, сумма ширин которых равна 100, *и круг радиуса* 1. *Докажите, что каждую из полос можно параллельно перенести так, чтобы все они покрывали круг.*

Разберём решение этой задачи, предложенное в статье [15]. Рассмотрим выпуклую фигуру F, граница которой состоит из кривой AB и отрезков AO и BO. Существуют два вида полосок: имеющие ограниченное и неограниченное пересечение с углом AOB (см. рис. 20).

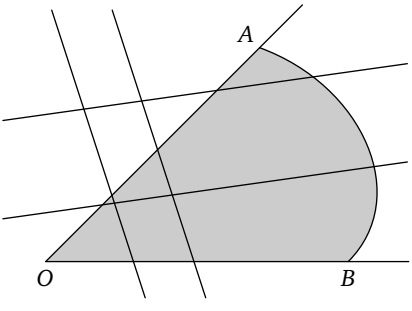


Рис. 20

Пусть есть полоски, имеющие неограниченное пересечение с углом AOB, причём прямые, ограничивающие каждую

полоску, не параллельны OA и OB. Покажем индукцией по числу полосок, что если сумма их ширин не меньше длины кривой AB, то F можно покрыть сдвигами этих полосок. Для одной полоски утверждение очевидно. Если есть несколько параллельных друг другу полосок, объединим их в одну полоску, ширина которой равна суммарной ширине этих полосок. Упорядочим направления полосок против часовой стрелки и возьмём крайнее из направлений. Перенесём соответствующую полоску так, чтобы одна из ограничивающих её прямых касалась нашей фигуры и при этом полоска имела общие точки с фигурой (возможны два случая взаимного расположения фигуры F и полоски, см. рис. 21).

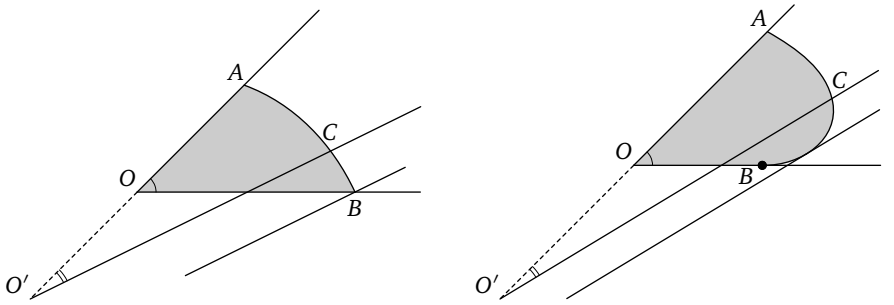


Рис. 21

Пусть вторая ограничивающая полоску прямая пересекается с кривой AB в точке C, а с прямой OA — в точке O'. От кривой AB будет отрезана кривая CB, длина которой не меньше, чем ширина полоски. Оставшимися полосками покроем фигуру, ограниченную отрезками $O'A$, $O'C$ и кривой AC (это можно сделать по предположению индукции).

Применим доказанное утверждение к фигуре, изображённой на рис. 22. В качестве кривой возьмём верхнюю полуокружность и два вертикальных отрезка AC и BD, в качестве угла возьмём развёрнутый угол AOB (прямая AB выбрана таким образом, что она не параллельна ни одной из полосок). Получаем, что если сумма ширин полосок хотя бы $\pi + 2$, то круг можно покрыть параллельными переносами полосок.

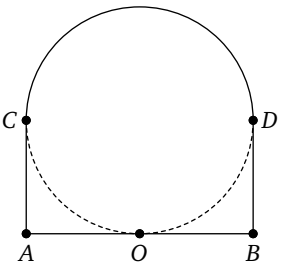


Рис. 22

Можно ли ещё уменьшить суммарную ширину полосок, чтобы утверждение задачи осталось верным? Ответ на этот вопрос неизвестен. Пользуясь леммой Банга, несложно построить пример полосок, суммарная ширина которых сколь угодно близка к π, но покрыть единичный круг их параллельными переносами нельзя.

Рассмотрим вписанный в единичный круг правильный $2n$-угольник. Пусть O — центр круга, a — длина стороны $2n$-угольника. Каждой паре параллельных сторон $2n$-угольника сопоставим перпендикулярную этим сторонам полоску, ширина которой равна a/λ для некоторого $\lambda > 1$ (пример такой полоски — на рис. 23). Обозначим через $\vec{v}_1, \ldots, \vec{v}_n$ векторы, каждый из которых параллелен соответствующей стороне $2n$-угольника (из двух возможных направлений выбирается произвольное) и по длине равен $a/2$. Тогда векторы $\vec{v}_1/\lambda, \ldots, \vec{v}_n/\lambda$ перпендикулярны прямым, ограничивающим соответствующие полоски, и по длине равны половинам ширин соответствующих полосок. По лемме Банга множество точек

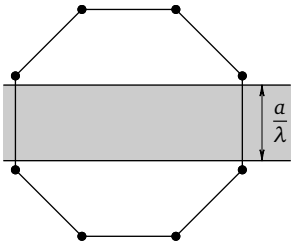


Рис. 23

$$O + \lambda \cdot \left(\pm\frac{\vec{v}_1}{\lambda} \pm \frac{\vec{v}_2}{\lambda} \pm \ldots \pm \frac{\vec{v}_n}{\lambda} \right) = O \pm \vec{v}_1 \pm \vec{v}_2 \pm \ldots \pm \vec{v}_n$$

не покрыто полосками. Докажем, что все 2^n точек лежат внутри исходного $2n$-угольника.

Рассмотрим одну из двух сторон, соответствующих вектору $\vec{v}_1$. Обозначим её через A_1A_2, а параллельную ей сторону $2n$-угольника — через $A_{n+1}A_{n+2}$ (см. рис. 24). Докажем, что каждая из 2^n точек либо находится в той же полуплоскости относительно прямой A_1A_2, что и точка O, либо лежит на прямой A_1A_2. Рассмотрим вектор $\overrightarrow{OX}$, где X — середина A_1A_2. Возьмём такие $\alpha_2, \alpha_3, \ldots, \alpha_n \in \{-1, 1\}$, что угол между каждым из векторов $\alpha_2\vec{v}_2$, $\alpha_3\vec{v}_3, \ldots, \alpha_n\vec{v}_n$ и вектором $\overrightarrow{OX}$ острый. Тогда

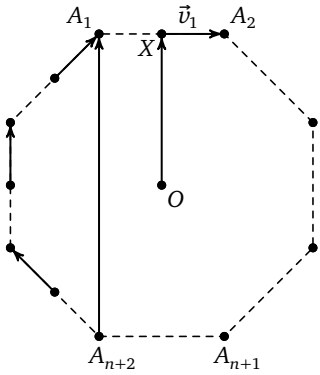


Рис. 24

$$\alpha_2\vec{v}_2 + \alpha_3\vec{v}_3 + \ldots + \alpha_n\vec{v}_n = \frac{1}{2}\overrightarrow{A_{n+2}A_1} = \overrightarrow{OX},$$

поэтому пара точек

$$O \pm \vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3 + \ldots + \alpha_n\vec{v}_n$$

совпадает с парой точек A_1, A_2. Очевидно, что, если изменить знаки перед какими-нибудь из векторов $\vec{v}_2, \ldots, \vec{v}_n$ на противоположные, полученные точки будут лежать в той же полуплоскости относительно A_1A_2, что и точка O.

Проведя аналогичное рассуждение для любой стороны $2n$-угольника, получим, что все 2^n точек лежат внутри или на границе многоугольника, а следовательно, и внутри круга. Таким образом, круг не покрыт полосками целиком. Поскольку при стремлении n к бесконечности периметр пра-

вильного $2n$-угольника стремится к 2π, сумма ширин полосок может быть сколь угодно близка к π за счёт выбора n и λ, откуда следует требуемое.

В случае произвольной выпуклой фигуры в работе [12] получены следующие оценки на суммарную ширину полосок, при которых выпуклую фигуру гарантированно можно покрыть параллельными переносами этих полосок. Пусть p и ω — периметр и ширина выпуклой фигуры F соответственно, D — её диаметр (т. е. наибольшее из расстояний между точками фигуры). Пусть также на плоскости расположены n полосок с ширинами $\omega_1, \omega_2, \ldots, \omega_n$. Тогда если выполнено хотя бы одно из условий

$$\omega_1 + \omega_2 + \ldots + \omega_n \geqslant \frac{3}{\pi} p,$$
$$\omega_1 + \omega_2 + \ldots + \omega_n \geqslant 2\sqrt{2} D,$$
$$\omega_1 + \omega_2 + \ldots + \omega_n \geqslant D + 2\omega,$$

то фигуру F можно покрыть параллельными переносами полосок.

Для случая пространства неизвестно, всегда ли можно покрыть единичный шар параллельными переносами слоёв, если суммарная ширина слоёв бесконечна. Наилучший результат для трёхмерного случая можно найти в работе [14]. Здесь мы приведём лишь наглядный частный случай этого результата, тем не менее сохранив основную идею доказательства.

Теорема. *Если в пространстве расположены слои $S_1, S_2, \ldots$, причём ширина S_i равна $\omega_i = 1/i$ для всех i, то их параллельными переносами можно покрыть всё пространство.*

Доказательство. Докажем, что для любого k существует такое N_k, что параллельными переносами слоёв $S_k, S_{k+1}, \ldots, S_{N_k}$ можно покрыть шар F_0 некоторого фиксированного радиуса $\varepsilon > 0$ (мы проведём доказательство для $\varepsilon = 1/32$). Из этого будет следовать утверждение теоремы. Действительно, разобьём все слои на множества

$$\widetilde{S}_j = \{S_{i_j+1}, S_{i_j+2}, \ldots, S_{i_{j+1}}\}, \quad j = 1, 2, \ldots,$$

каждым из которых можно покрыть шар радиуса ε. Поскольку шарами радиуса ε можно покрыть всё пространство, то и параллельными переносами слоёв можно покрыть всё пространство.

Каждому слою S_i сопоставим слой S'_i с той же срединной плоскостью, но вдвое меньшей ширины. Обозначим некоторый шар радиуса $\frac{1}{32}$ через F_0. Опишем «жадный» алгоритм, следуя которому можно покрыть бóльшую часть объёма F_0 с помощью параллельного переноса слоёв $S'_k, S'_{k+1}, \ldots$

Для каждого $i = 1, 2, \ldots$ обозначим через F_i часть шара, оставшуюся не покрытой параллельными переносами слоёв $S'_k, S'_{k+1}, \ldots, S'_{k+i-1}$. Пока-

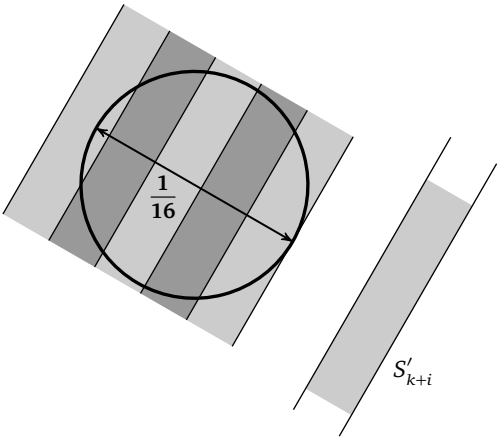


Рис. 25

жем, как по F_i строить F_{i+1}. Заметим, что F_i можно покрыть с помощью

$$\left\lceil \frac{1/16}{\omega_{k+i}/2} \right\rceil = \left\lceil \frac{k+i}{8} \right\rceil$$

слоёв, каждый из которых является параллельным переносом слоя S'_{k+i} (см. рис. 25). Хотя бы один из этих слоёв покрывает

$$\frac{1}{\lceil (k+i)/8 \rceil} > \frac{8}{k+i+8}$$

от объёма F_i. В этот слой и перенесём S'_{k+i}, получив F_{i+1}. Тогда для всех i будет выполнено неравенство

$$\mathrm{Vol}(F_i) < \mathrm{Vol}(F_{i-1}) \cdot \Big(1 - \frac{8}{k+i+8}\Big).$$

Оценим объём F_i:

$$\begin{aligned}
\mathrm{Vol}(F_i) &< \mathrm{Vol}(F_{i-1}) \cdot \Big(1 - \frac{8}{k+i+8}\Big) < \\
&< \mathrm{Vol}(F_{i-2}) \cdot \Big(1 - \frac{8}{k+i+8}\Big) \cdot \Big(1 - \frac{8}{k+(i-1)+8}\Big) < \ldots < \\
&< \mathrm{Vol}(F_0) \cdot \prod_{j=1}^{i} \Big(1 - \frac{8}{k+j+8}\Big) = \mathrm{Vol}(F_0) \cdot \exp\Big(\sum_{j=1}^{i} \ln\Big(1 - \frac{8}{k+j+8}\Big)\Big).
\end{aligned}$$

Поскольку $\ln(1-x) < -x$ при $x \in (0;1)$, имеем

$$\mathrm{Vol}(F_0) \cdot \exp\Big(\sum_{j=1}^{i} \ln\Big(1 - \frac{8}{k+j+8}\Big)\Big) < \mathrm{Vol}(F_0) \cdot \exp\Big(-8 \sum_{j=1}^{i} \frac{1}{k+j+8}\Big).$$

Как известно, предел $\lim_{n\to\infty}\left(\sum_{i=1}^{n}\frac{1}{n}-\ln n\right)$ конечен: он носит название постоянной Эйлера — Маскерони и равен 0,57... Следовательно, существует такое i, начиная с которого выполнено неравенство

$$\sum_{j=1}^{i}\frac{1}{k+j+8}>\frac{1}{2}\ln i.$$

Таким образом, для этого i

$$\mathrm{Vol}(F_i)<\mathrm{Vol}(F_0)\cdot e^{-4\ln i}=\frac{\pi}{3\cdot 2^{13}}\cdot\frac{1}{i^4}.$$

При достаточно большом i объём $\mathrm{Vol}(F_i)$ будет меньше объёма шара с радиусом $\frac{1}{2(k+i)}$, поэтому в F_i нельзя будет поместить шар такого радиуса. Значит, каждая точка в F_i удалена от какого-то из слоёв $S'_1,\ldots,S'_i$ на расстояние, меньшее $\frac{1}{2(k+i)}$. Поэтому если рассмотреть вместо слоёв $S'_1,\ldots,S'_i$ слои $S_1,\ldots,S_i$, получим, что они целиком покрывают F_0. Тогда в качестве N_k можно взять это значение i. Теорема доказана. □

Отметим, что использованный при доказательстве теоремы подход, а именно сжатие тела в некоторое число раз и последующее применение жадного или подобного алгоритма, неоднократно встречается в задачах об упаковках и покрытиях. По всей вероятности, впервые эта идея была предложена в работе [9].

Список литературы

[1] *Смуров М.*, *Спивак А.* Покрытие полосками // Квант. 1998. № 4. С. 17–22.

[2] *Яглом И. М.* Т. Банг — В. Фенхель. Решение одной задачи о покрытии выпуклых фигур // Математическое просвещение. Сер. 2. Вып. 1. М.: Гостехиздат, 1957. С. 214–218.

[3] *Alexander R.* A problem about lines and ovals // Amer. Math. Monthly. 1968. Vol. 75, №. 5. P. 482–487.

[4] *Ball K.* The plank problem for symmetric bodies // Invent. Math. 1991. Vol. 104, № 3. P. 535–543.

[5] *Bang T.* A solution of the “plank problem” // Proc. Amer. Math. Soc. 1951. Vol. 2, № 6. P. 990–993.

[6] *Bezdek A.* On a generalization of Tarski’s plank problem // Discrete Comput. Geom. 2007. Vol. 38, № 2. P. 189–200.

[7] *Bognár M.* On W. Fenchel’s solution of the plank problem // Acta Math. Acad. Sci. Hungar. 1961. Vol. 12, № 3–4. P. 269–270.

[8] *Davenport H.* A note on Diophantine approximation // Studies in mathematical analysis and related topics. Stanford, Calif: Stanford Univ. Press, 1962. P. 77–81.

[9] *Erdös P., Rogers C. A.* Covering space with convex bodies // Acta Arithm. 1962. Vol. 7. P. 281–285.

[10] *Green J. W.* On the determination of a function in the plain by its integrals over straight lines // Proc. Amer. Math. Soc. 1958. Vol. 9. P. 758–762.

[11] *Groemer H.* On coverings of plane convex sets by translates of strips // Aequationes Math. 1981. Vol. 22, № 2–3. P. 215–222.

[12] *Groemer H.* Some remarks on translative coverings of convex domains by strips // Canad. Math. Bull. Vol. 27, №. 2. 1984. P. 233–237.

[13] *Kadets V.* Coverings by convex bodies and inscribed balls // Proc. Amer. Math. Soc. 2005. Vol. 133, № 5. P. 1491–1495.

[14] *Kupavskii A., Pach J.* From Tarski's plank problem to simultaneous approximation // Amer. Math. Monthly. 2017. Vol. 124, № 6. P. 494–505.

[15] *Makai E., Pach J.* Controlling function classes and covering Euclidean space // Stud. Scient. Math. Hungarica. 1983. Vol. 18, № 2–4. P. 435–459.

[16] *Moese H.* Przyczynek do problemu A. Tarskiego: «O stopniu równoważności wielokątów» (English: A contribution to the problem of A. Tarski "On the degree of equivalence of polygons") // Parametr. 1931–1932. Vol. 2. P. 305–309.

[17] *Tarski A.* O stopniu równoważności wielokątów (English: On the degree of equivalence of polygons) // Mlody Matematyk. 1931. Vol. 1. P. 37–44.

[18] *Tarski A.* Uwagi o stopniu równoważności wielokątów (English: Remarks on the degree of equivalence of polygons) // Parametr. 1932. Vol. 2. P. 310–314.

Алексей Вадимович Доледенок, Центр педагогического мастерства (г. Москва)

doledenok@gmail.com

Анна Николаевна Доледенок, МГУ имени М. В. Ломоносова

anya11235@mail.ru

Наш семинар: математические сюжеты

Среднее число случайных слагаемых в растущей сумме, достигшей заданного значения

И. Р. Высоцкий

Сумма независимых одинаковых случайных величин, имеющих дискретное или непрерывное равномерное распределение на каком-то промежутке, распределена по закону, близкому к нормальному. Это факт хорошо известен. Если дано число слагаемых, то указать интервал, в который попала такая сумма с любой наперёд заданной вероятностью, несложно. Ошибка при этом обусловлена заменой истинного распределения нормальным, и, как правило, ничтожно мала даже при не очень большом числе слагаемых.

Интересно поставить обратную задачу: каково должно быть число случайных слагаемых определённого вида, чтобы их сумма *впервые* достигла некоторого наперёд заданного значения.

Мы начнём издалека. Сначала будем бросать обычную игральную кость и складывать выпадающие очки. Нас будет интересовать вероятность того, что в какой-то момент сумма выпавших очков станет равна некоторому числу n, а также математическое ожидание числа сделанных бросков к моменту, когда число n будет достигнуто.

Затем перейдём к непрерывному случаю, складывая независимые случайные величины, равномерно распределённые на интервале $(0; 1)$.

Основная задача — найти математическое ожидание числа слагаемых в такой сумме в момент, когда она стала больше или равна наперёд задан-

ного числа x. А именно, мы покажем, что это математическое ожидание равно

$$m(x) = \sum_{j=0}^{[x]} \frac{(-1)^j (x-j)^j e^{x-j}}{j!}$$

и что на бесконечности функция $m(x)$ асимптотически приближается к линейной функции $y = 2x + 2/3$.

§ 1. Стоп-числа в игре с правильной костью

Все в детстве любили (многие любят и во взрослом состоянии) незамысловатые игры: на столе картонное поле с извилистой тропинкой. Игроки по очереди бросают кость. Сколько очков выпало, на столько шагов игрок продвигает свою фишку, с которой по дороге случаются разные приятные и неприятные события. Фишка может отъехать назад, попав на несчастливое поле, или наоборот — проскочить несколько полей, заработать для хозяина бросок вне очереди и т. п.

Предположим, что мы играем в такую игру, но без счастливых или несчастливых полей. Самый простой и естественный вопрос: с какой вероятностью фишка в какой-то момент остановится на поле с номером n? Все ли поля равновероятны? Если кость симметричная, то задача легко формализуется.

Задача 1. Какова вероятность, что при последовательных бросаниях правильной игральной кости сумма очков в какой-то момент станет в точности равна n?

Решение. Назовём *стоп-числом* число, которое в какой-то момент стало суммой. Интуиция подсказывает, что поскольку каждый бросок увеличивает сумму в среднем на 3,5, то стоп-числами окажутся в среднем 2 из каждых 7 чисел. Это рассуждение и даёт результат, весьма близкий к точному.

Дадим точное решение задачи, приводящее к рекуррентной формуле. Рассмотрим по очереди возможности, возникающие после первого броска. Если первый бросок дал 1, то n будет стоп-числом, только если $n-1$ тоже стоп-число. Так же обстоит дело со всеми исходами первого броска: если первый бросок дал k очков ($k = 1, \ldots, 6$), то n будет стоп-числом, только если $n-k$ также стоп-число, образовавшееся при последующих бросках. Обозначим через p_n вероятность того, что n — стоп-число. Тогда

$$p_n = \frac{1}{6} \cdot p_{n-1} + \frac{1}{6} \cdot p_{n-2} + \ldots + \frac{1}{6} \cdot p_{n-6} = \frac{1}{6} \sum_{k=1}^{6} p_{n-k}. \qquad (1)$$

Из свойств линейной рекурсии известно, что последовательность p_n представима в виде суммы геометрических прогрессий со знаменателями, модуль которых не превосходит 1, при этом хотя бы один из них в точности равен 1 (поскольку сумма коэффициентов в правой части равна 1). Поэтому рекурсия имеет пределом некоторое число, которое, как мы понимаем, должно равняться $2/7$. Получить явное выражение для p_n непросто, даже если удастся решить соответствующее характеристическое уравнение 6-й степени.

Чтобы понять, как ведут себя вероятности p_n, зададим начальные условия: $p_{-5} = \ldots = p_{-1} = 0$, $p_0 = 1$. Это естественное предположение: отрицательной сумма быть не может, а нулевой она является с вероятностью $p_0 = 1$ до того, как кость брошена первый раз. Тогда

$$p_1 = \frac{1}{6} = \frac{7^0}{6}, \quad p_2 = \frac{1}{6}\Big(1 + \frac{1}{6}\Big) = \frac{1}{6} \cdot \frac{7}{6} = \frac{7^1}{6^2}$$

и так далее: при $n \leqslant 6$

$$p_n = \frac{1}{6}\Big(\frac{7^{n-2}}{6^{n-1}} + \frac{7^{n-3}}{6^{n-2}} + \ldots + \frac{7^0}{6} + 1\Big) = \frac{1}{6} \cdot \Big(\frac{1}{6} \cdot \frac{(7/6)^{n-1} - 1}{7/6 - 1} + 1\Big) = \frac{7^{n-1}}{6^n}.$$

Такая закономерность сохраняется, пока последняя ненулевая вероятность p_{n-6} в сумме (1) равна единице. Таким образом, поначалу вероятности p_n образуют растущую геометрическую прогрессию, которая достигает наибольшего значения при $n = 6$. При $n > 6$ эта закономерность нарушается, поскольку при $n = 7$ в сумме (1) последнее слагаемое $p_0 = 1$ уступает место слагаемому $p_1 = 1/6$. Затем снова рост до $n = 11$. Далее происходят затухающие колебания вероятностей с периодом 5–6 шагов. Локальные максимумы вероятностей выделены в таблице 1 жирным шрифтом, при этом разумно считать, что уже при $n \geqslant 20$ вероятность p_n неотличима от $2/7$.

Если вы составляете детскую игру с кубиком и фишками (см. с. 104) и хотите сделать её повеселее, устройте какую-нибудь первую неприятность (или наоборот) как раз на поле 6.

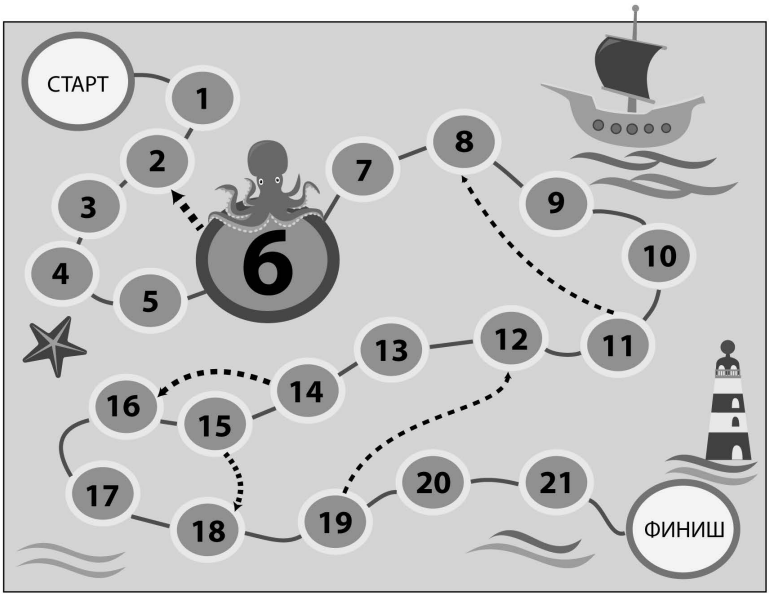
Интересно посмотреть на «трудозатраты», нужные для того, чтобы достичь поля n. В случае с обычным кубиком для этого потребуется от $n/6$ бросков (если невероятно повезёт) до n бросков (при столь же невероятном невезении).

Задача 2. Найти математическое ожидание числа бросков, сделанных к моменту, когда сумма очков впервые оказалась больше или равна n.

Решение. Пусть случайная величина S_n — число бросков, которые пришлось сделать, чтобы сумма очков достигла n. Введём для чисел

Таблица 1

n	p_n	n	p_n	n	p_n	n	p_n
−5	0,000000000000	15	0,286113932137	35	0,285710193751	55	0,285714292240
−4	0,000000000000	**16**	**0,287071429920**	36	0,285711733841	56	0,285714283629
−3	0,000000000000	17	0,286701924733	37	0,285715051280	57	0,285714280890
−2	0,000000000000	18	0,285586725149	**38**	**0,285716315054**	58	0,285714284162
−1	0,000000000000	19	0,284712810463	39	0,285714800621	59	0,285714287275
0	**1,000000000000**	20	0,285621080152	40	0,285713653567	**60**	**0,285714287468**
1	0,166666666667	**21**	**0,285967983759**	41	0,285713624686	61	0,285714285944
2	0,194444444444	22	0,285943659029	42	0,285714196508	62	0,285714284895
3	0,226851851852	23	0,285755697214	**43**	**0,285714606953**	63	0,285714285106
4	0,264660493827	24	0,285597992628	44	0,285714532898	64	0,285714285808
5	0,308770576132	25	0,285599870541	45	0,285714235872	**65**	**0,285714286083**
6	**0,360232338820**	26	0,285747713887	46	0,285714141747	66	0,285714285884
7	0,253604395290	**27**	**0,285768819510**	47	0,285714223111	67	0,285714285620
8	0,268094016728	28	0,285735625468	48	0,285714322848	68	0,285714285566
9	0,280368945441	29	0,285700953208	**49**	**0,285714343905**	69	0,285714285678
10	0,289288461040	30	0,285691829207	50	0,285714300064	**70**	**0,285714285773**
11	**0,293393122242**	31	0,285707468637	51	0,285714261258	71	0,285714285767
12	0,290830213260	**32**	**0,285725401653**	52	0,285714265489	72	0,285714285715
13	0,279263192334	33	0,285721682947	53	0,285714286112	73	0,285714285686
14	0,283539658507	34	0,285713826853	**54**	**0,285714296613**	74	0,285714285697



$k = 1, \ldots, 6$ шесть бинарных случайных величин I_k — индикаторов событий «первый бросок дал ровно k очков»:

$$I_k = \begin{cases} 1, & \text{если первый бросок дал ровно } k \text{ очков,} \\ 0, & \text{если первый бросок дал не } k \text{ очков.} \end{cases}$$

Тогда

$$S_n = I_1 S_{n-1} + I_2 S_{n-2} + \ldots + I_6 S_{n-6} + 1,$$

где случайная величина S_{n-k} означает число бросков, не считая первого, потребовавшихся для того, чтобы достичь суммы очков $n - k$. Единица в конце нужна, чтобы учесть первый бросок. Величины I_k и S_{n-k} независимы, поскольку индикатор I_k относится только к первому броску. Введём для ожиданий $\mathrm{E}S_n$ краткое обозначение m_n, перейдём в полученном равенстве к математическому ожиданию и получим линейное неоднородное рекуррентное уравнение

$$m_n = 1 + \frac{1}{6} \sum_{k=1}^{6} m_{n-k}, \tag{2}$$

отличающееся от (1) только ненулевым свободным членом. Неудивительно, что частным решением уравнения (2) является последовательность $m_n = \frac{2}{7} n$ ($n \in \mathbb{N}$) — это легко проверить непосредственно. Общее решение (2) имеет вид

$$m_n = \frac{2}{7} n + a_n,$$

где a_n — некоторая сходящаяся последовательность, которая даёт решение соответствующего однородного уравнения; она зависит от начальных условий.

Взяв для рекурсии (2) естественные начальные условия $m_n = 0$ при $n \leqslant 0$ и $m_1 = 1$ (сумма достигнет единицы обязательно при первом же броске), можно провести расчёт. Он показывает, что

$$m_n \approx \frac{2}{7} n + \frac{10}{21}.$$

Если кость брошена ровно 10 раз, то ожидание суммы выпавших очков равно 35. Обратим ситуацию (см. таблицу 2) и видим, что для достижения суммы 35, требуется в среднем не 10 бросков, а 10 «с хвостиком»: больше на примерно 10/21. Природу хвостика понять нетрудно: 35 окажется стоп-числом с вероятностью около 2/7. А с вероятностью 5/7 сумма очков перескочит число 35, не остановившись на нём. Вот на это перескакивание и расходуются лишние примерно $\frac{10}{21}$ броска.

Таблица 2

n	m_n	Отличие m_n от $\frac{2}{7}n+\frac{10}{21}$
32	9,6190338922	−0,00001373
33	9,9047592939	−0,00000261
34	10,1904809768	0,00000479
35	**10,4761948037**	**0,00000433**
36	10,7619049974	0,00000024
37	11,0476167313	−0,00000232
38	11,3333317826	−0,00000155

§ 2. Непрерывный случай. Точное решение задачи

Если слагаемые — не очки на кубике, а непрерывные случайные величины, то ситуация должна быть примерно такой же, как в случае с кубиком: математическое ожидание числа слагаемых должно приблизительно линейно зависеть от наперёд заданного стоп-числа x. При этом тоже должен быть какой-то «хвостик», но, вероятно, уже не 10/21, а какой-то другой. Можно поэкспериментировать на компьютере, бросая разные n-гранные кости и увеличивая n, но мы попробуем решить задачу в общем виде. Рассмотрим теперь последовательность одинаковых и независимых случайных величин ξ_i $(i=1,2,\ldots)$, равномерно распределённых на интервале $(0;1)$.

Задача 3. Найти математическое ожидание числа независимых равномерно распределённых на интервале $(0;1)$ величин ξ_i к моменту, когда их сумма впервые достигла числа x.

Обозначим через $S(x)$ или просто S число слагаемых в сумме

$$\tau=\tau_S=\xi_1+\xi_2+\ldots+\xi_S,$$

которая впервые достигла числа x. Очевидно, $S(x)>x$. Величина S является случайной функцией аргумента x, поскольку каждое x порождает распределение $p_s=\mathrm{P}(S=s)=\mathrm{P}(\tau_{s-1}<x\leqslant\tau_S)$. Природа вероятностей p_s известна — они выражаются через плотности f_s распределений[1)] сумм τ_s:

$$p_s(x)=\int\limits_{x-1}^{x} f_{s-1}(t)(t+1-x)\,dt=\int\limits_{0}^{1} t f_{s-1}(t+x-1)\,dt \qquad (3)$$

[1)] Распределения Ирвинга — Холла. Функции f_s являются многочленами.

(для общности положим $f_0(x)$ равной дельта-функции $\delta(x)$). Используя соотношение (3), видимо, можно получить вероятности p_s в явном виде. Но мы применим для поиска $\mathrm{E}S(x)$ иной метод.

Для краткости записи введём функцию $m(x) = \mathrm{E}S(x)$, которая непрерывна при $x > 0$. Это интуитивно ясно — при малом изменении x дискретная целочисленная величина S с вероятностью, стремящейся к единице, не изменяется. Строгое доказательство непрерывности функции $m(x)$ при $x > 0$ можно провести непосредственно — это несложное упражнение в математическом анализе. Другой путь: непрерывность следует из того, что функции f_s и, следовательно, вероятности p_s непрерывны.

Вопрос о математическом ожидании случайной величины S приводит к непрерывному аналогу уравнения (2):

$$\mathrm{E}S = m(x) = 1 + \int_0^1 m(x-t)\,dt. \tag{4}$$

Получился частный (для равномерно распределённых слагаемых) случай уравнения, которое называют *уравнением восстановления*[2)]. Причину такого названия обсудим ниже. Перепишем уравнение иначе:

$$m(x) = 1 + \int_{x-1}^{x} m(t)\,dt. \tag{5}$$

При $x < 0$ следует считать, что $m(x) = 0$. При этом можно доопределить функцию в нуле: из уравнения (5) получается

$$m(0) = 1 + \int_{-1}^{0} m(t)\,dt = 1.$$

Это согласуется с соображением, что требуется одно слагаемое, чтобы сумма стала положительной, т. е. $\lim_{x\to 0+} m(x) = 1$. Приняв эти соглашения, получим начальные условия:

$$\begin{cases} m(x) = 0, & \text{если } x < 0, \\ m(0) = 1. \end{cases} \tag{6}$$

Таким образом, мы ищем непрерывное при $x > 0$ решение задачи (5)–(6).

Если функция $y = m(x)$ непрерывна в точках x и $x-1$, то из (5) следует:

$$m'(x) = m(x) - m(x-1), \tag{7}$$

[2)] Общий вид и вывод уравнения восстановления можно найти, например, в статье [5]. См. также [4, 6].

т. е. функция $m(x)$ дифференцируема в точке x. Значит, функция $m(x)$ дифференцируема во всех точках, кроме точки 0 (где она имеет разрыв) и точки 1 (поскольку имеет разрыв в точке 0).

Сделаем небольшое техническое упрощение — введём вспомогательную функцию $g(x) = e^{-x}m(x)$. Тогда уравнение (7) и условия (6) преобразуются в задачу

$$g'(x) = -\frac{1}{e}g(x-1), \quad \begin{cases} g(x) = 0 & \text{при } x < 0, \\ g(0) = 1. \end{cases} \tag{8}$$

Уравнение (8), впрочем, как и уравнение (7), — линейное дифференциальное уравнение с *запаздывающим аргументом*[3)].

Обычно такие уравнения решаются по шагам на последовательных промежутках, длина которых равна запаздыванию. Сначала найдём решение на промежутке $[0; 1)$, потом на промежутке $[1; 2)$ и так далее. На каждом промежутке получается обычная задача Коши, поскольку решение на предыдущем промежутке уже известно. На $[0; 1)$ получаем задачу

$$g'(x) = 0, \quad g(0) = 1,$$

откуда $g(x) = 1$.

Должно выполняться равенство

$$g(1) = \lim_{x \to 1-} g(x) = 1.$$

Поэтому на промежутке $[1; 2)$ получается задача

$$g'(x) = -\frac{1}{e}, \quad g(1) = 1,$$

имеющая единственное решение $g(x) = 1 - (x-1)e^{-1}$ при $1 \leqslant x < 2$.

Продвигаясь таким образом далее, получаем:

$$g(x) = 1 - \frac{x-1}{e} + \frac{(x-2)^2}{2e^2} - \frac{(x-3)^3}{6e^3} + \ldots + \frac{(x-j)^j}{j!\,e^j}$$

при $j \leqslant x < j+1$. Тогда

$$m(x) = \sum_{j=0}^{[x]} \frac{(-1)^j (x-j)^j e^{x-j}}{j!}. \tag{9}$$

Интересно посмотреть на результат при всех натуральных $x = n \in \mathbb{N}$. Формула (9) принимает вид

$$m(n) = \sum_{j=0}^{n} \frac{(-1)^j (n-j)^j}{j!} e^{n-j} = \sum_{k=0}^{n} \frac{(-1)^{n-k} k^{n-k}}{(n-k)!} e^k. \tag{10}$$

[3)] См., например, неоднократно переизданную монографию [3].

Получается некоторый многочлен степени n, вычисленный в точке e. В частности, чтобы достичь единицы, требуется $m(1) = e$ слагаемых; чтобы сумма достигла двух, потребуется в среднем

$$m(2) = e^2 - (2-1)e^{2-1} = e^2 - e$$

слагаемых и так далее:

$$m(3) = e^3 - 2e^2 + \frac{1}{2}e, \quad m(4) = e^4 - 3e^2 + 2e - \frac{1}{6}e, \quad \dots$$

Результат удобно проиллюстрировать таблицей, в которую столбиками выпишем члены рядов для последовательных отрицательных степеней e: e^{-1}, e^{-2}, e^{-3}, … При этом каждый следующий столбик сместим на одну строку вниз по отношению к предыдущему. Во всех столбцах каждое число, кроме верхней единицы, получается умножением сверху стоящего числа на множитель $-k/(n-k)$.

x \ e^{-k}	e^{-1}	e^{-2}	e^{-3}	e^{-4}	e^{-5}	e^{-6}	e^{-7}	…	e^{-k}	…
1	1									
2	-1	1								
3	$1/2$	-2	1							
4	$-1/6$	2	-3	1						
5	$1/24$	$-4/3$	$9/2$	-4	1					
6	$-1/120$	$2/3$	$-9/2$	8	-5	1				
7	$1/720$	$-4/15$	$27/8$	$-32/3$	$25/2$	-6	1			
…	…	…	…	…	…	…	…	…		
n	$\frac{(-1)^{n-1}}{(n-1)!}$	$\frac{(-2)^{n-2}}{(n-2)!}$	$\frac{(-3)^{n-3}}{(n-3)!}$	…	…	…	…	…	$\frac{(-k)^{n-k}}{(n-k)!}$	…
…	…	…	…	…	…	…	…	…	…	…

Сумма чисел в k-м столбце равна e^{-k}. Если умножить числа k-го столбца на e^k, сумма в каждом столбце станет равна 1, а сумма чисел в n-й строке окажется равна $\mathrm{E}S(n)$. Приведём результаты с точностью до четырёх знаков в таблице 3.

Суммы быстро приближаются к $2n + 2/3$. Возникает предположение, что

$$m(x) = 2x + \frac{2}{3} + o(1). \tag{11}$$

Таблица 3

Степени e^k	2,7183	7,3891	20,0855	54,5982	148,4132	403,429	1096,63	2980,96	8103,08	22026,5	ES: сумма чисел строки, умноженных на соответствующие степени e^k
n \ k	1	2	3	4	5	6	7	8	9	10	
1	1										**2,7183**
2	−1	1									**4,6708**
3	0,5	−2	1								**6,6666**
4	−0,1667	2	−3	1							**8,6667**
5	0,0417	−1,3333	4,5	−4	1						**10,6667**
6	−0,0083	0,6667	−4,5	8	−5	1					**12,6667**
7	0,0014	−0,2667	3,375	−10,6667	12,5	−6	1				**14,6667**
8	−0,0002	0,0889	−2,025	10,6667	−20,8333	18	−7	1			**16,6667**
9	0,0002	−0,0254	1,0125	−8,5333	26,0417	−36	24,5	−8	1		**18,6667**
10	−0,0000	0,0063	−0,4339	5,6889	−26,0417	54	−57,1667	32	−9	1	**20,6667**
	.....										

§ 3. Асимптотическая оценка

Докажем равенство (11). Из общей теории линейных уравнений с запаздывающим аргументом следует, что непрерывное при $x > 0$ решение задачи (6)–(7) приближается к некоторой линейной функции. Если подставить частное решение $m = ax + b$ в уравнение (5), то несложно найти, что $a = 2$, т. е. решение (9) с ростом x асимптотически приближается к линейной функции $y = 2x + b$. Обычно формулируют более слабое утверждение, которое в наших терминах имеет вид

$$\frac{m(x)}{x} \xrightarrow[x\to\infty]{} \frac{1}{\mathrm{E}\xi_1} = 2,$$

и получается из простой и важной леммы Вальда (Wald's equation), которую мы докажем позже.

Этот результат согласуется с естественным интуитивным предположением, что увеличение суммы на 1 требует в среднем двух дополнительных слагаемых, поскольку каждое в среднем равно 0,5.

На рис. 1 показан график функции $y = m(x)$. Уже при $2 < x \leqslant 3$ график функции $m(x)$ невозможно на глаз отличить от прямой.

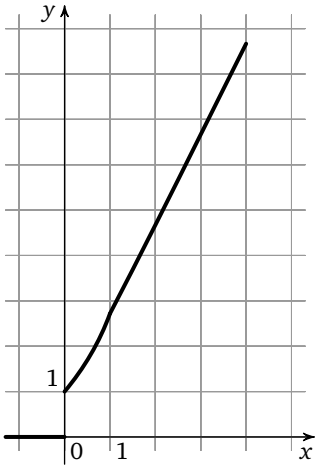


Рис. 1

Пока что решённая задача относится к тем, где точное решение, доставив эстетическое удовольствие, не снимает главный вопрос: а всё же, сколько это будет? Расчёт по формуле (9) затруднителен, а асимптотика

$$m(x) \sim 2x + \frac{2}{3}$$

требует доказательства, хотя ясно видна в таблице 3.

Сначала докажем две леммы, причём первая потребуется позже, когда мы будем обсуждать связь нашей задачи с числами Эйлера.

Лемма 1 о случайном остатке. *Пусть η — произвольная действительная случайная величина, а $\xi \sim U(0;1)$ — стандартная равномерно распределённая случайная величина, и эти величины независимы. Тогда дробная часть суммы $\{\eta + \xi\}$ равномерно распределена на промежутке $[0;1)$, при этом величины $\{\eta + \xi\}$ и η независимы.*

Доказательство. Нужно показать, что на промежутке $[0;1)$ функция распределения суммы $\eta + \xi$ тождественно равна x: $F_{\eta+\xi}(x) = x$ независимо от распределения величины η. Не ограничивая общности, можно

считать, что $0 \leqslant \eta < 1$. Рассмотрим два случая: $0 \leqslant x \leqslant \eta$ и $\eta < x < 1$:

$$F_{\eta+\xi}(x) = \mathrm{P}(\eta + \xi \leqslant x) =$$
$$= \begin{cases} \mathrm{P}(1 \leqslant \eta + \xi < x + 1), & \text{если } 0 \leqslant x < \eta, \\ \mathrm{P}(\eta \leqslant \eta + \xi < x) + \mathrm{P}(1 \leqslant \eta + \xi < \eta + 1), & \text{если } \eta \leqslant x < 1 \end{cases} =$$
$$= \begin{cases} \mathrm{P}(1 - \eta \leqslant \xi < x + 1 - \eta), & \text{если } 0 \leqslant x < \eta, \\ \mathrm{P}(0 \leqslant \xi < x - \eta) + \mathrm{P}(1 - \eta \leqslant \xi < 1), & \text{если } \eta \leqslant x < 1 \end{cases} =$$
$$= \begin{cases} x, & \text{если } 0 \leqslant x < \eta, \\ x - \eta + \eta, & \text{если } \eta \leqslant x < 1 \end{cases} = x.$$

Отсюда следует утверждение леммы. □

Иллюстрируют эту лемму стрелочные часы: если кто-то в какой-то момент случайным образом покрутит минутную стрелку, то после этого часы будут показывать совершенно случайное время безо всякой связи с тем, что они показывали до того.

Если имеется дополнительная информация о слагаемых, то ситуация меняется. Например, если известно, что $\eta < 0$ и $\eta + \xi \geqslant 0$, то сумма $\eta + \xi$ может иметь уже вовсе не равномерное распределение.

Лемма 2 о случайном остатке. *Пусть случайные величины ξ и η независимы, причём величина ξ равномерно распределена на интервале $(0; 1)$, а величина η равномерно распределена на интервале $(-1; 0)$. Тогда*

$$\mathrm{E}(\xi + \eta \mid \xi + \eta \geqslant 0) = \frac{1}{3}.$$

Доказательство. На координатной плоскости $\xi O \eta$ условия

$$T = \{0 < \xi < 1,\ -1 < \eta < 0,\ \xi + \eta \geqslant 0\}$$

определяют треугольник (см. рис. 2), на котором равномерно распределён случайный вектор (ξ, η). Поэтому

$$\mathrm{E}(\xi + \eta \mid T) = \frac{\iint\limits_T (x + y)\, dx\, dy}{\iint\limits_T dx\, dy}.$$

Добавим третью координатную ось (см. рис. 3). Числитель дроби равен объёму пирамиды с основанием T и с вершиной в точке $(1; 0; 1)$, а знаменатель равен площади треугольника T. Поэтому

$$\mathrm{E}(\xi + \eta \mid T) = \frac{1/6}{1/2} = \frac{1}{3}.$$

Лемма 2 доказана. □

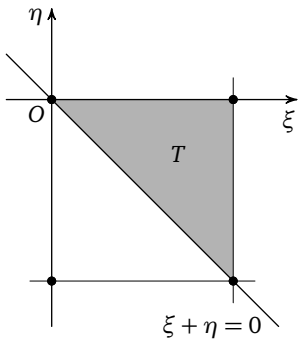


Рис. 2

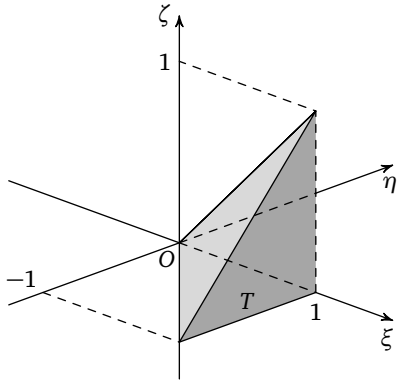


Рис. 3

Вернёмся к суммам равномерных слагаемых и математическому ожиданию их числа. Пусть $x > 0$. Рассмотрим последовательность независимых случайных величин $\xi_1, \xi_2, \xi_3, \ldots, \xi_k, \ldots$, имеющих стандартное равномерное распределение $U(0,1)$. Будем составлять суммы

$$\tau_1 = \xi_1, \quad \tau_2 = \xi_1 + \xi_2, \quad \tau_s = \tau_{s-1} + \xi_s, \quad \ldots$$

до тех пор, пока очередная сумма τ не окажется больше или равна числа x. Как и прежде, число слагаемых в сумме τ обозначим S:

$$\tau = \tau_S = \sum_{k=1}^{S} \xi_k.$$

Сперва докажем ещё одну важную лемму, которую ещё называют леммой или равенством Вальда[4] [5].

Лемма 3. $\mathrm{E}\tau = \mathrm{E}\xi_1 \cdot \mathrm{E}S$ (*в нашем случае* $\mathrm{E}\tau = \frac{1}{2}\mathrm{E}S$).

Доказательство. Введём индикаторы I_s событий «$s-1$ слагаемого оказалось недостаточно для достижения суммы x» при каждом натуральном $s > 0$:

$$I_s = \begin{cases} 0, & \text{если } \tau_{s-1} \geqslant x, \\ 1, & \text{если } \tau_{s-1} < x. \end{cases}$$

Тогда $S = I_1 + I_2 + \ldots$ и $\tau = I_1\xi_1 + I_2\xi_2 + \ldots$

Ясно, что $I_1 = 1$ и что каждый следующий индикатор I_s зависит только от τ_{s-1}, но не зависит от ξ_s. Поэтому, переходя к математическим ожида-

[4] «Мы знаем, что это равенство важное, поскольку носит чьё-то имя», — пишет Петер Небрес в своей статье [5]. Присоединимся к его мнению.

ниям, получим:

$$\mathrm{E}\tau = \sum_{s=1}^{\infty} \mathrm{E}(I_s \xi_s) = \sum_{s=1}^{\infty} \mathrm{E}I_s \mathrm{E}\xi_s = \mathrm{E}\xi_1 \sum_{s=1}^{\infty} \mathrm{E}I_s = \frac{1}{2} \sum_{s=1}^{\infty} \mathrm{E}I_s = \frac{1}{2}\mathrm{E} \sum_{s=1}^{\infty} I_s = \frac{1}{2}\mathrm{E}S. \qquad \square$$

Лемма Вальда выявляет связь между $\mathrm{E}S$ и $\mathrm{E}\tau$, а лемма 2 будет нужна при доказательстве равенства $\mathrm{E}\tau = x + 1/3 + o(1)$. Собрав эти результаты воедино, мы получим $\mathrm{E}S = 2x + 2/3 + o(1)$.

Чтобы получить равенство $\mathrm{E}\tau = x + 1/3 + o(1)$, достаточно убедиться, что в сумме

$$\tau = \left(\sum_{k=1}^{S-1} \xi_k - x \right) + \xi_S + x$$

слагаемое, взятое в скобки, имеет асимптотически равномерное распределение на $(-1; 0)$:

$$\sum_{k=1}^{S-1} \xi_k - x \xrightarrow[x \to \infty]{} \eta,$$

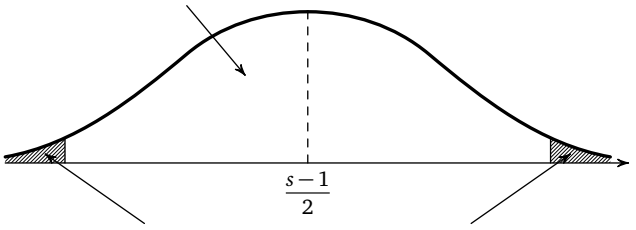
где η — некоторая случайная величина, имеющая равномерное распределение $U(-1, 0)$. Это, в свою очередь, следует из того, что сумма

$$\tau_{s-1} = \sum_{k=1}^{s-1} \xi_k$$

имеет распределение, очень близкое к нормальному $N\left(\frac{s-1}{2}, \frac{s-1}{12}\right)$, которое в своей средней части, будучи сужено на любой единичный интервал, с ростом s приближается к равномерному на этом интервале (см. рис. 4).

На хвостах нормального распределения это не так, но поведение на хвостах не важно, поскольку при подходящем выборе границ средней части

Здесь, в средней части, нормальное распределение с ростом s приближается к равномерному на любом единичном интервале



А вероятность того, что τ_{s-1} окажется с краю, исчезающе мала

Рис. 4

вероятность того, что τ_{s-1} окажется на каком-то из хвостов, стремительно приближается к 0 при росте x, а значит, при росте s, поскольку $s > x$.

Проведём эти рассуждения теперь аккуратно и в правильном порядке.

Утверждение. $\mathrm{E}\tau = 2x + 2/3 + o(1)$ *при* $x \to \infty$.

Доказательство. Центрируя и нормируя случайную величину τ_{s-1}, получаем случайную величину

$$\zeta_{s-1} = \frac{\tau_{s-1} - (s-1)/2}{\sqrt{(s-1)/12}} = 2\sqrt{3}\,\frac{\tau_{s-1} - (s-1)/2}{\sqrt{s-1}},$$

значения которой при условии $x - 1 < \tau_{s-1} < x$ принадлежат интервалу

$$A = \Big(2\sqrt{3}\,\frac{x-1-(s-1)/2}{\sqrt{s-1}};\ 2\sqrt{3}\,\frac{x-(s-1)/2}{\sqrt{s-1}}\Big).$$

Длина интервала A равна $\dfrac{2\sqrt{3}}{\sqrt{s-1}}$. Увеличивая x, мы тем самым увеличиваем s, и последовательность величин ζ_{s-1} сходится к стандартной нормальной случайной величине, а её плотности $f_{\zeta_{s-1}}(t)$ равномерно сходятся к стандартной нормальной плотности

$$\varphi(t) = \frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{t^2}{2}\Big).$$

Разобьём $\mathrm{E}\tau$ в сумму условных математических ожиданий при условиях

$$H = \{A \subset (-s^{1/4}; s^{1/4})\} \quad \text{и} \quad \overline{H} = \{A \not\subset (-s^{1/4}; s^{1/4})\}.$$

Получим:

$$\mathrm{E}\tau = \mathrm{E}(\tau|H)\mathrm{P}(H) + \mathrm{E}(\tau|\overline{H})\mathrm{P}(\overline{H}).$$

Поскольку $s \xrightarrow[x\to\infty]{} \infty$, получаем $\mathrm{P}(H) \xrightarrow[x\to\infty]{} 1$ и $\mathrm{P}(\overline{H}) \xrightarrow[x\to\infty]{} 0$. Поэтому

$$\mathrm{E}(\tau|H) - \mathrm{E}\tau \xrightarrow[x\to\infty]{} 0, \tag{12}$$

и вместо $\mathrm{E}\tau$ можно искать $\mathrm{E}(\tau|H)$, ограничиваясь гипотезой H. При этом условии рассмотрим относительный размах функции плотности $f_{\zeta_{s-1}}$ на интервале A:

$$\frac{\sup_A f_{\zeta_{s-1}} - \inf_A f_{\zeta_{s-1}}}{\inf_A f_{\zeta_{s-1}}} = \frac{\sup_A f_{\zeta_{s-1}}}{\inf_A f_{\zeta_{s-1}}} - 1.$$

При достаточно большом x можно заменить в этом выражении сложно устроенную плотность $f_{\zeta_{s-1}}$ стандартной нормальной плотностью φ (в силу

равномерной сходимости f_{ζ_s} к φ), при этом расширив для удобства интервал A до его замыкания $[A]$:

$$\frac{\sup_A f_{\zeta_{s-1}}}{\inf_A f_{\zeta_{s-1}}} - 1 + o(1) = \frac{\max_{[A]} \varphi}{\min_{[A]} \varphi} - 1 = \frac{\exp(-t_1^2/2)}{\exp(-t_2^2/2)} - 1 = \exp\left(\frac{t_2^2 - t_1^2}{2}\right) - 1,$$

где t_1 и t_2 — точки, в которых φ принимает соответственно наибольшее и наименьшее значения на отрезке $[A]$. Очевидно,

$$|t_1|, |t_2| \leqslant s^{1/4} \quad \text{и} \quad |t_2 - t_1| \leqslant \frac{2\sqrt{3}}{\sqrt{s-1}}.$$

Тогда

$$\exp\left(-\frac{2\sqrt{3}\, s^{1/4}}{\sqrt{s-1}}\right) - 1 \leqslant \frac{\max \varphi}{\min \varphi} - 1 = \frac{\exp(-t_1^2/2)}{\exp(-t_2^2/2)} - 1 \leqslant \exp\left(\frac{2\sqrt{3}\, s^{1/4}}{\sqrt{s-1}}\right) - 1.$$

Левая и правая части этого неравенства стремятся к 0 при $s \to \infty$. Следовательно,

$$\frac{\max_{[A]} \varphi}{\min_{[A]} \varphi} - 1 \xrightarrow[x \to \infty]{} 0$$

и поэтому

$$\frac{\sup_A f_{\zeta_{s-1}} - \inf_A f_{\zeta_{s-1}}}{\inf_A f_{\zeta_{s-1}}} \xrightarrow[x \to \infty]{} 0.$$

Таким образом, какое бы значение ни приняла случайная величина $S > x$, с ростом x распределение случайной величины

$$\tau - \xi_S = \sum_{k=1}^{S-1} \xi_k$$

приближается к равномерному на интервале $(x-1; x)$, а случайная величина $\tau - \xi_S - x$ приближается к случайной величине η, равномерно распределённой на интервале $(-1; 0)$. С помощью леммы 2 получаем:

$$\begin{aligned} \mathrm{E}(\tau | H, \tau \geqslant x) &= \mathrm{E}(\xi_S + (\tau - \xi_S - x) \mid \tau \geqslant x) + x + o(1) = \\ &= \mathrm{E}(\xi_S + \eta \mid \xi_S + \eta \geqslant 0) + x + o(1) = x + \frac{1}{3} + o(1). \end{aligned}$$

Учитывая (12), видим, что

$$\mathrm{E}\tau = x + \frac{1}{3} + o(1).$$

Отсюда и из леммы 3 следует:

$$\mathrm{E}S = 2x + \frac{2}{3} + o(1).$$

Чтобы сумма независимых стандартных равномерных слагаемых достигла числа x, в среднем потребуется приблизительно $2x + 2/3$ слагаемых. Добиться с помощью случайных слагаемых суммы, в точности равной x, практически невозможно. Почти наверняка сумма превзойдёт число x на величину, в среднем равную примерно $1/3$. Вот на эту «лишнюю» треть и расходуются в среднем лишних $2/3$ случайного слагаемого. □

Рассмотренная задача — частный случай одной из основных задач так называемой *теории восстановления* (Renewal Theory), о которой в российской математической литературе практически нет упоминаний. Известно переводное издание [2], адресованное преимущественно инженерам и, вероятно, поэтому вышедшее в издательстве «Советское радио».

В общем случае под процессом обновления или восстановления (Renewal Process) имеется в виду случайный процесс, состоящий из последовательных сумм независимых неотрицательных случайных величин, имеющих какой-то определённый смысл. Процесс продолжается до тех пор, пока сумма этих величин не достигнет наперёд заданного значения.

Классический пример — работа прибора, который должен функционировать до определённого момента (например, до окончания гарантийного срока или до выработки ресурса). Однако прибор может ломаться и до истечения гарантии, и в таких случаях подлежит ремонту (восстановлению). Случайные слагаемые здесь — периоды времени между поломками, сумма которых даёт полное время работы прибора.

Возникают естественные вопросы — каково среднее количество этих случайных слагаемых, т. е. поломок, случившихся к моменту окончания гарантийного срока, и во сколько в среднем это обойдётся. Более серьёзный вопрос — не следует ли в какой-то момент из соображений экономии заменить исправный прибор новым ещё до окончания ресурса?

Из предыдущего абзаца ясно, что породило математическую теорию восстановления. Исследователей в приложениях больше интересовали процессы восстановления, где каждый интервал между двумя поломками имеет показательное или близкое к показательному распределение, что породило представление о процессе восстановления как об обобщении пуассоновского процесса.

§ 4. Связь задачи о числе слагаемых с числами Эйлера

Подойдём к рассмотренной задаче с комбинаторной стороны. Правда, придётся пожертвовать произвольностью достигаемой суммы. Поставим вопрос только о целых суммах и временно забудем всё, сделанное прежде.

Начнём с простого случая $n = 1$: сколько в среднем случайных слагаемых нужно, чтобы достичь единицы? Этот случай поможет понять смысл и логику общего построения. Возьмём случайное слагаемое ξ_1 и прибавим к нему ξ_2. Если сумма меньше единицы, добавим ξ_3, и так до тех пор, пока полученная сумма не достигнет $n = 1$. Присмотримся к остаткам — дробные части последовательно получающихся сумм τ_s равны

$$Z_1 = \{\tau_1\} = \{\xi_1\} = \xi_1, \quad Z_2 = \{\tau_1 + \xi_2\}, \quad Z_3 = \{\tau_2 + \xi_3\}, \quad \ldots$$

Найдём вероятность события $S > s$, т. е. события $\tau_s < 1$ ($m \in \mathbb{N}$).

Во-первых, в силу леммы 1 все остатки Z_i равномерно распределены на интервале $(0; 1)$: $Z_i \sim U(0, 1)$, во-вторых, все они независимы попарно и в совокупности. В-третьих, остатки растут, пока $\tau_s < 1$, но как только сумма достигает 1, очередной остаток оказывается меньше предыдущего. Таким образом, событие $\tau_{s-1} < 1 \leqslant \tau_s$ случается только тогда, когда с остатками происходит событие $Z_1 < Z_2 < Z_3 < \ldots < Z_{s-1} > Z_s$, а событие $\tau_s < 1$ эквивалентно событию $Z_1 < Z_2 < Z_3 < \ldots < Z_{s-1} < Z_s$.

Нам удалось переформулировать задачу: имеются s последовательных независимых случайных чисел, равномерно распределённых на интервале $(0; 1)$, и нужно найти вероятность того, что они случайным образом расположились по возрастанию в порядке их появления. Вероятность этого, очевидно, равна $1/s!$.

Для события $\tau_s < 1$ введём индикатор

$$I_s = \begin{cases} 1, & \text{если } \tau_s < 1, \\ & \text{т. е. если } s \text{ слагаемых не хватило для достижения суммы } n = 1, \\ 0, & \text{если } \tau_s \geqslant 1. \end{cases}$$

Случайная величина S легко выражается через эти индикаторы:

$$S = 1 + I_1 + I_2 + I_3 + \ldots = 1 + \sum_{s=1}^{\infty} I_s.$$

В этой сумме каждый следующий индикатор I_s равен единице, если s слагаемых не хватило, чтобы сумма достигла числа 1. Как только сумма 1 достигнута, все последующие индикаторы равны нулю.

Учитывая, что

$$\mathrm{E}I_s = \mathrm{P}(I_s = 1) = \mathrm{P}(\tau_s < 1) = \frac{1}{s!},$$

и переходя к математическим ожиданиям, получаем:

$$\mathrm{E}S = 1 + \sum_{s=1}^{\infty} \mathrm{E}I_s = 1 + \sum_{s=1}^{\infty} \frac{1}{s!} = \sum_{s=0}^{\infty} \frac{1}{s!} = e.$$

Получился уже известный результат: среднее количество слагаемых, необходимых для достижения в сумме единицы, равно e.

Возможно обобщение на другие целые суммы. Принцип тот же. Сумма n будет впервые достигнута при s слагаемых, если $\tau_{s-1} < n \leqslant \tau_s$. Так же как раньше, найдём вероятности события «s слагаемых недостаточно для достижения суммы n», т. е. события $\tau_s < n$. Это событие эквивалентно тому, что в последовательности независимых остатков $Z_1, Z_2, \ldots, Z_s$ ровно $n-1$ число меньше предыдущего. Заменив случайные числа Z_i их *рангами* $\rho(i)$ (ранг — номер в упорядоченном ряду), вместо случайной последовательности остатков получаем случайную перестановку их рангов. Таким образом, задача сводится к вопросу о том, какова вероятность того, что в случайной перестановке $(\rho(1), \rho(2), \rho(3), \ldots, \rho(s))$ длины s наблюдается ровно $n-1$ *падение*, т. е. ровно $n-1$ пара $\rho(i)$, $\rho(i+1)$, где $\rho(i) > \rho(i+1)$. Ответ на этот вопрос дают *числа Эйлера*[5], см., например, [1, с. 297–300].

Определение. Числом Эйлера I рода $A(s,i)$ называется количество перестановок длины s, которые содержат ровно i падений (или подъёмов).

Очевидно, $A(s,0) = 1$ при $s > 0$ — этим равенством мы пользовались в случае $n = 1$. Для общности обычно полагают $A(0,0) = 1$, $A(s,i) = 0$ при $i < 0$ или при $s < 0$, а также при $0 < s \leqslant i$. При $s > 0$, $i \geqslant 0$ справедливо основное рекуррентное соотношение

$$A(s,i) = (i+1)A(s-1,i) + (s-i)A(s-1,i-1).$$

Имеется и явная формула:

$$A(s,i) = \sum_{j=0}^{i} (-1)^j C_{s+1}^j (i+1-j)^s.$$

Вывод этих соотношений остаётся за рамками статьи.

С помощью чисел Эйлера можно записать вероятность события $A_s < n$:

$$\mathrm{P}(A_s < n) = \sum_{i=0}^{n-1} \frac{A(s,i)}{s!}.$$

Снова введём индикаторы

$$I_s = \begin{cases} 1, & \text{если } A_s < n, \\ 0, & \text{если } A_s \geqslant n, \end{cases}$$

[5] Числа Эйлера (или числа Эйлера I рода) не следует путать с эйлеровыми числами, которые возникают при разложении в степенной ряд гиперболического секанса.

составим сумму

$$S = 1 + \sum_{s=1}^{\infty} I_s$$

и перейдём к математическим ожиданиям:

$$\mathrm{E}S = 1 + \sum_{m=1}^{\infty} \mathrm{P}(I_m = 1) = 1 + \sum_{m=1}^{\infty} \mathrm{P}(A_m < n) =$$
$$= 1 + \sum_{m=1}^{\infty} \sum_{i=0}^{n-1} \frac{A(m,i)}{m!} = \sum_{m=0}^{\infty} \sum_{i=0}^{n-1} \frac{A(m,i)}{m!}. \qquad (13)$$

Последнее преобразование сделано с учётом соглашения

$$A(0,0) = 1, \quad A(0,i) = 0 \quad \text{при } i > 0.$$

Читатель может ради любопытства и удовольствия самостоятельно из равенства (13) получить равенство (10). Потребуются несколько утомительные, но естественные комбинаторные преобразования.

Список литературы

[1] *Кнут Д.*, *Грэхем Р.*, *Паташник О.* Конкретная математика. Основание информатики. М.: Мир, 1998.

[2] *Кокс Д. Р.*, *Смит В. Л.* Теория восстановления. М.: Советское радио, 1967.

[3] *Мышкис А. Д.* Линейные дифференциальные уравнения с запаздывающим аргументом. М.: Наука, 1972.

[4] *Doob J. L.* Renewal theory from the point of view of the theory of probability // Trans. AMS. 1948. Vol. 63. P. 422–438.

[5] *Nebres P.* Renewal theory and its applications. The University of Chicago, 2011. http://math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/Nebres.pdf

[6] *Ross Sh. M.* The inspection paradox // Probab. Engrg. Inform. Sci. 2003. Vol. 17, № 1. P. 47–51.

Иван Ростиславович Высоцкий, МЦНМО
i_r_vysotsky@hotmail.com

О вычислении классических сумм Якобсталя

Н. Н. Осипов

В статье рассказывается о быстрых алгоритмах вычисления классических сумм Якобсталя. Эти алгоритмы основаны на нетривиальной связи сумм Якобсталя с представлением простых чисел $p \equiv 1 \pmod 4$ и $p \equiv 1 \pmod 3$ бинарными квадратичными формами $A^2 + B^2$ и $A^2 + 3B^2$ соответственно.

Введение

Пусть $p > 2$ — нечётное простое число. Сумма вида

$$\phi(n) = \sum_{x=0}^{p-1} \left(\frac{x^3 + nx}{p} \right) \tag{0.1}$$

называется *суммой Якобсталя* (она впервые возникла и изучалась в работе Э. Якобсталя 1907 года [21]). Здесь n — произвольное целое число, а $\left(\frac{a}{p}\right)$ обозначает *символ Лежандра* a по p: если $a \equiv 0 \pmod p$, то $\left(\frac{a}{p}\right) = 0$, иначе

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{если сравнение } x^2 \equiv a \pmod p \text{ разрешимо,} \\ -1 & \text{в противном случае.} \end{cases}$$

Основные свойства символа Лежандра излагаются практически в любом учебнике по элементарной теории чисел (см., например, [10]). В частности, имеет место сравнение

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod p$$

(так называемый *критерий Эйлера*), с помощью которого можно быстро вычислить символ Лежандра $\left(\frac{a}{p}\right)$ даже при больших p[1)].

[1)] Речь идёт о применении хорошо известного *бинарного алгоритма* возведения в степень [34]. Более эффективный алгоритм основан на замене символа Лежандра его обобщением — *символом Якоби*, который быстро вычисляется с помощью *квадратичного закона взаимности* (см. [10, гл. 6]).

Многие авторы изучали и более общие суммы вида

$$\phi_l(n)=\sum_{x=0}^{p-1}\Big(\frac{x^{l+1}+nx}{p}\Big),\quad l\in\mathbb{N},\tag{0.2}$$

которые также называются суммами Якобсталя (подробности см. в статье [30] или, начиная с п. 5.49, в хорошо известной монографии [9], где основательно освещена и история вопроса).

В данной статье будет рассказано, как вычисляются суммы Якобсталя в случаях $l=2$ и $l=3$[2)]. При $n\not\equiv 0 \pmod p$ имеем

$$\phi_3(n)=\sum_{x=1}^{p-1}\Big(\frac{x^4+nx}{p}\Big)=\sum_{x=1}^{p-1}\Big(\frac{x^{-4}+nx^{-1}}{p}\Big)=\sum_{x=1}^{p-1}\Big(\frac{1+nx^3}{p}\Big)=\Big(\frac{n}{p}\Big)\sum_{x=0}^{p-1}\Big(\frac{x^3+n^{-1}}{p}\Big),$$

где $^{-1}$ означает взятие обратного по модулю p. Вместо суммы $\phi_3(n)$ (которая впервые, по-видимому, изучалась в работе [27]) нам будет удобнее рассматривать сумму

$$\psi(n)=\sum_{x=0}^{p-1}\Big(\frac{x^3+n}{p}\Big).\tag{0.3}$$

Суммы Якобсталя (0.1) и (0.3) имеют непосредственное отношение к следующей важной задаче (в том числе для приложений в криптографии [7, гл. VI]). Рассмотрим *эллиптическую кривую* E, заданную уравнением

$$y^2=x^3+ax+b,\tag{0.4}$$

над полем $\mathbb{Z}_p$ классов вычетов по модулю p (предполагается, что p не является делителем дискриминанта $\Delta=-4a^3-27b^2$). Как известно, на множестве $\mathbb{Z}_p$-точек (x,y) этой кривой вместе с формальной бесконечно удалённой точкой ∞ можно ввести операцию сложения, относительно которой это множество превращается в *абелеву группу* (при этом ∞ играет роль нулевого элемента). Порядок $N_p(E)$ этой группы выражается формулой

$$N_p(E)=\sum_{x=0}^{p-1}\Big(1+\Big(\frac{x^3+ax+b}{p}\Big)\Big)+1=p+1+\sum_{x=0}^{p-1}\Big(\frac{x^3+ax+b}{p}\Big).$$

Согласно *теореме Хассе* о числе точек эллиптической кривой над конечным полем, имеет место неравенство $|N_p(E)-p-1|<2p^{1/2}$ (теорема 10.5 в книге [6]). Нахождение точного значения $N_p(E)$ при больших p является содержательной задачей, для решения которой существует довольно

[2)] Случай $l=1$ совсем прост (см. далее лемму 1.5).

нетривиальный *алгоритм Шуфа* (см. оригинальную работу [26]; описание алгоритма можно быстро найти по ссылкам [33, 35]). В частном случае, когда $b = 0$ или $a = 0$, имеется более простой способ найти значение $N_p(E)$, поскольку для сумм $\phi(n)$ и $\psi(n)$, как выясняется, есть быстрый практический алгоритм вычисления[3)].

Цель настоящей статьи — рассказать о практически пригодных (быстро работающих даже для больших p) алгоритмах вычисления сумм Якобсталя $\phi(n)$ и $\psi(n)$ при любом n. С теоретической основой этих алгоритмов можно познакомиться по книгам [1] (см. §§ 18.3, 18.4) и [13] (см. теоремы 6.2.9 и 6.2.10). Описание самих алгоритмов с иллюстрирующими примерами можно найти, например, в статье [22]. Для доказательства утверждений, на которых основаны алгоритмы, обычно привлекают мощный аппарат *сумм Гаусса* и *сумм Якоби*, но мы воспользуемся более элементарными средствами в духе идей пионерской работы [21].

§ 1. Предварительные сведения

В этом разделе мы сообщим предварительные сведения, необходимые для дальнейшего. Те утверждения, которые можно быстро (и не сильно отвлекаясь) доказать, мы приведём с доказательством, а в остальном ограничимся комментариями и ссылками на соответствующую литературу.

1.1. Алгоритм Корнаккиа

Ключевым моментом в задаче быстрого вычисления классических сумм Якобсталя $\phi(n)$ и $\psi(n)$ является неожиданная и нетривиальная связь с задачей представления простого числа p в виде $p = A^2 + B^2$ (для суммы $\phi(n)$) или $p = A^2 + 3B^2$ (для суммы $\psi(n)$). Последняя задача интересна и сама по себе, но для нас принципиально то, что она имеет быстрый алгоритм решения, причём даже в следующей более общей постановке.

Пусть d — фиксированное натуральное число. Требуется представить простое число $p > d$ в виде

$$p = A^2 + dB^2, \tag{1.1}$$

где A и B — некоторые натуральные числа. Конечно, для данного d представление (1.1) возможно не для любого простого числа p, так как есть необходимое условие — разрешимость сравнения

$$x^2 + d \equiv 0 \pmod{p}. \tag{1.2}$$

[3)] В случае суммы (0.1) см. иллюстрирующий пример 1.4 в книге [23, гл. 6].

Последнее легко выяснить с помощью квадратичного закона взаимности. Например, при $d = 1$ сравнение (1.2) разрешимо тогда и только тогда, когда $p \equiv 1 \pmod 4$. Вместе с тем, дать несложное описание тех простых p, которые можно представить в виде (1.1), удаётся далеко не для всех d (см. по этому поводу книгу [18]).

В интересующих нас случаях $d = 1$ и $d = 3$ такое описание есть. Для $d = 1$ его доставляет знаменитая *теорема Ферма — Эйлера*, которая гласит: всякое простое число $p \equiv 1 \pmod 4$ допускает представление в виде суммы двух квадратов. Для $d = 3$ имеет место аналогичное утверждение: каждое простое число $p \equiv 1 \pmod 3$ допускает представление в виде суммы квадрата и утроенного квадрата. Оба утверждения могут быть легко получены на основе *факториальности* колец целых чисел соответствующих мнимых квадратичных полей (см., например, доказательство теоремы Ферма — Эйлера в учебнике [8, гл. 4, § 2, с. 153]; вообще, эту классическую теорему можно доказать многими способами, среди которых есть и весьма экзотические [31]).

Для сравнения рассмотрим $d = 11$. Легко проверить, что для простого $p > 11$ сравнение (1.2) разрешимо тогда и только тогда, когда $p \equiv 1, 3, 4, 5, 9 \pmod{11}$. Но последнее условие ещё не гарантирует возможность представления (1.1): например, для $p = 23$ оно выполнено, однако равенство $23 = A^2 + 11B^2$ невозможно. Какие же простые числа $p > 11$ можно представить в виде $p = A^2 + 11B^2$? Известен только такой ответ: те и только те, для которых сравнение

$$x^3 - x^2 - x - 1 \equiv 0 \pmod p$$

имеет три решения. Этот критерий весьма нетривиален, но он не упрощает проверку[4)].

Прежде чем переходить к изложению *алгоритма Корнаккиа* (см. п. 1.5.2 в книге [17]), решающего для данного простого p вопрос о представлении в виде (1.1), докажем следующий важный факт.

Лемма 1.1. *Если представление* (1.1) *возможно, то оно однозначно*[5)].

Доказательство. Будем рассуждать от противного. Пусть

$$p = A^2 + dB^2 = X^2 + dY^2,$$

где A, B, X, Y — натуральные числа, причём $A < X$, $B > Y$. Поскольку

$$B^2X^2 - A^2Y^2 \equiv B^2X^2 + dB^2Y^2 = B^2(X^2 + dY^2) = B^2p \equiv 0 \pmod p,$$

имеем одно из двух сравнений

$$BX \pm AY \equiv 0 \pmod p.$$

[4)] Подробности и переформулировку критерия в терминах *чисел Трибоначчи* см. в [19].

[5)] В случае $d = 1$ мы не обращаем внимание на порядок слагаемых.

Так как $BX \pm AY > 0$, получаем $BX \pm AY \geqslant p$. Следовательно,

$$p^2 \leqslant (BX \pm AY)^2 \leqslant (A^2 + B^2)(X^2 + Y^2) \leqslant p^2,$$

причём равенство возможно только при $d=1$ и в этом случае вектор (X, Y) пропорционален одному из векторов $(B, \pm A)$, откуда $X = B$ и $Y = A$. □

Для успешной работы алгоритма Корнаккиа требуется предварительно найти некоторое решение $x = m_0$ сравнения (1.2). В случае больших простых p это тоже содержательная задача, для решения которой можно применить общий *вероятностный алгоритм* решения уравнений над полем $\mathbb{Z}_p$ или специфические методы извлечения квадратных корней по модулю p (см., например, §§ 6.1, 6.2 в книге [2]). Вероятностные алгоритмы такого рода на практике работают довольно быстро.

Пусть $\left(-\frac{d}{p}\right) = 1$, т. е. сравнение (1.2) разрешимо. Тогда для $p = 4k + 3$ его решения можно найти по явной формуле $x \equiv \pm(-d)^{k+1} \pmod p$, поскольку

$$x^2 \equiv (-d)^{2k+2} = (-d)^{(p+1)/2} \equiv -d \pmod p.$$

В случае $p \equiv 1 \pmod 4$ можно применить *алгоритм Тонелли — Шенкса* [36], для чего предварительно потребуется найти какой-нибудь *квадратичный невычет* b по модулю p. На практике можно просто «подбросить монетку»: выбрать случайный вычет b и вычислить символ Лежандра $\left(\frac{b}{p}\right)$: с вероятностью $1/2$ он будет равен -1. В случаях $p \equiv 5 \pmod 8$ и $p \equiv 5 \pmod{12}$ можно взять $b = 2$ и $b = 3$ соответственно.

Сравнение (1.2) можно также решить с помощью *алгоритма Чиполлы* [32]. В этом алгоритме вычисления производятся в некотором *квадратичном расширении* поля $\mathbb{Z}_p$. Предварительно необходимо найти вычет b, для которого $b^2 + 4d$ является квадратичным невычетом по модулю p (понятно, что и здесь применимы вероятностные соображения).

Будем считать, что $0 < m_0 < p/2$. Алгоритм Корнаккиа состоит в следующем.

(а) Положим $r_0 = p$, $r_1 = m_0$.

(б) Применяя *алгоритм Евклида* к числам r_0 и r_1, будем вычислять остатки $r_2, \ldots, r_s$ (r_{i+1} — остаток от деления r_{i-1} на r_i) до тех пор, пока не получим неравенство $r_s^2 < p$.

(в) Если $(p - r_s^2)/d = t^2$ для некоторого натурального t, то (1.1) имеет место для $(A, B) = (r_s, t)$. Иначе представление (1.1) невозможно.

В случае $d = 1$ шаг (в) можно упростить: вычислим ещё один остаток r_{s+1}, и тогда $(A, B) = (r_s, r_{s+1})$. На практике алгоритм Корнаккиа довольно быстро решает вопрос о представлении (1.1), поскольку последовательность остатков $\{r_i\}$ экспоненциально убывает.

Пример 1.1. Пусть $d = 3$, $p = 2017$. Тогда $m_0 = 589$. Имеем

$$r_1 = p \bmod m_0 = 250, \quad r_2 = m_0 \bmod r_1 = 89, \quad r_3 = r_1 \bmod r_2 = 72,$$
$$r_4 = r_2 \bmod r_3 = 17, \quad 17^2 < 2017, \quad \frac{2017 - 17^2}{3} = 24^2.$$

Таким образом, $2017 = 17^2 + 3 \cdot 24^2$.

Обоснование корректности алгоритма Корнаккиа представляет интересную тему для отдельной статьи, и мы не будем здесь этим заниматься (читателя отсылаем к статье [12] как содержащей наиболее компактное изложение). Отметим только, что данный алгоритм работает[6)] и для составных p (при условии, что удалось найти все решения сравнения (1.2), а это сложная проблема в случае составного модуля).

Случай $d = 1$ издавна привлекал внимание. Известно несколько алгоритмов для представления простых чисел суммой двух квадратов, появившихся до алгоритма Корнаккиа (1908 год). В первую очередь следует упомянуть *алгоритм Эрмита — Серре* 1848 года (см. [20, 28]), использующий разложение рационального числа m_0/p в цепную дробь. По существу, алгоритм Корнаккиа в случае $d = 1$ является укороченным вариантом алгоритма Эрмита — Серре. На языке цепных дробей формулируется и схожий *алгоритм Смита* (1855 год), детальное изложение которого можно найти в статье [16]. Исторически первым был *алгоритм Лежандра* (1808 год), который использует разложение в (периодическую) цепную дробь квадратичной иррациональности $p^{1/2}$. Для больших p этот алгоритм, вообще говоря, неэффективен, так как период цепной дроби может оказаться настолько длинным, что его невозможно будет выписать.

Краткое изложение упомянутых алгоритмов (с поясняющими примерами) есть в главе V книги [5]. В статье [14] приводится обоснование усовершенствованной версии алгоритма Эрмита — Серре. Другое обоснование можно найти в статье [4].

1.2. Евклидовы кольца и алгоритм Евклида

Для некоторых d вопрос о получении представления (1.1) можно решать другим (тоже вполне эффективным на практике) способом. Нам понадобится понятие *евклидова кольца* (см., например, учебник [3]). Предполагая d свободным от квадратов, рассмотрим *кольцо целых чисел*

$$\mathbb{Z}[\omega] = \{x + y\omega : (x, y) \in \mathbb{Z}^2\}$$

[6)] В случае составного p алгоритм Корнаккиа находит только *примитивные представления* (1.1), т. е. с дополнительным условием НОД $(A, B) = 1$ (оно автоматически выполнено для простых p).

мнимого квадратичного поля $\mathbb{Q}(\sqrt{-d})$. Здесь

$$\omega = \begin{cases} \dfrac{1+\sqrt{-d}}{2} & \text{при } -d \equiv 1 \pmod 4, \\ \sqrt{-d} & \text{при } -d \not\equiv 1 \pmod 4. \end{cases}$$

Пусть $N(\gamma)$ — *норма* числа $\gamma = x + y\omega \in \mathbb{Q}(\sqrt{-d})$, т. е.

$$N(\gamma) = |\gamma|^2 = \gamma\overline{\gamma} = \begin{cases} \dfrac{x^2+xy+(d+1)y^2}{4} & \text{при } -d \equiv 1 \pmod 4, \\ x^2+dy^2 & \text{при } -d \not\equiv 1 \pmod 4 \end{cases}$$

(здесь и далее черта сверху означает сопряжение в поле комплексных чисел $\mathbb{C}$).

Следующий критерий очевидным образом вытекает из определений и свойства мультипликативности нормы: $N(\gamma_1\gamma_2) = N(\gamma_1)N(\gamma_2)$.

Лемма 1.2. *В кольце* $\mathbb{Z}[\omega]$ *можно делить с остатком относительно нормы* $N(\cdot)$ *тогда и только тогда, когда для любого* $\gamma \in \mathbb{Q}(\sqrt{-d})$ *найдётся такое* $\gamma^* \in \mathbb{Z}[\omega]$, *что*

$$N(\gamma - \gamma^*) < 1. \tag{1.3}$$

Опираясь на лемму 1.2, нетрудно установить евклидовость некоторых колец $\mathbb{Z}[\omega]$:

Лемма 1.3. *Кольцо* $\mathbb{Z}[\omega]$ *евклидово при* $d \in \{1, 2, 3, 7, 11\}$.

Доказательство. При естественном геометрическом изображении чисел кольца $\mathbb{Z}[\omega]$ на комплексной плоскости $\mathbb{C}$ получится *решётка* с базисом 1 и ω. Нужно убедиться, что базисный параллелограмм можно покрыть открытыми кругами единичного радиуса с центрами в вершинах 0, 1, ω и $1+\omega$ этого параллелограмма. Для указанных значений d это более или менее очевидно из элементарно-геометрических соображений. □

На самом деле в лемме 1.3 перечислены все евклидовы кольца: как бы мы ни вводили *евклидову норму* в кольце $\mathbb{Z}[\omega]$, других случаев евклидовости не появится (это несложное упражнение мы оставляем заинтересованному читателю). Случаи $d = 1$ (*целые гауссовы числа*) и $d = 3$ (*целые числа Эйзенштейна*) хорошо известны. Оценку (1.3) в этих случаях можно уточнить:

$$N(\gamma - \gamma^*) \leqslant \begin{cases} 1/2 & \text{при } d = 1, \\ 1/3 & \text{при } d = 3. \end{cases}$$

Для $d \in \{1, 2, 3, 7, 11\}$ задачу о представлении (1.1) можно решать следующим образом. Пусть простое число p таково, что сравнение (1.2)

разрешимо и $x = m_0$ — одно из решений. Идея состоит в том, чтобы рассмотреть $\delta = \text{НОД}\,(p, m_0 + \sqrt{-d})$ в кольце $\mathbb{Z}[\omega]$.

Лемма 1.4. $N(\delta) = p$.

Доказательство. Имеем

$$(m_0 + \sqrt{-d})(m_0 - \sqrt{-d}) = pl$$

для некоторого целого l. Отсюда видно, что $p > 2$ не является простым элементом евклидова (и потому факториального) кольца $\mathbb{Z}[\omega]$, ибо ни одно из чисел $m_0 \pm \sqrt{-d}$ не делится на p. Пусть $p = \pi\xi$, где $\pi,\ \xi \in \mathbb{Z}[\omega]$ и π — простой элемент. Тогда $p^2 = N(p) = N(\pi)N(\xi)$, откуда $N(\pi) = N(\xi) = p$ и $\xi = \overline{\pi}$. Таким образом, $p = \pi\overline{\pi}$ есть произведение двух простых (возможно, ассоциированных) элементов. Теперь уже легко установить, что δ ассоциировано либо с π, либо с $\overline{\pi}$. В обоих случаях $N(\delta) = p$. □

Теперь мы можем вычислить $\delta = x_0 + y_0\omega$ с помощью алгоритма Евклида для кольца $\mathbb{Z}[\omega]$ (это делается быстро даже для больших p). Тогда при $d \in \{1, 2\}$ имеем $p = x_0^2 + dy_0^2$, и задача о представлении (1.1) решена. Если же $d \in \{3, 7, 11\}$, то

$$p = x_0^2 + x_0y_0 + \frac{d+1}{4}y_0^2 = \left(x_0 + \frac{y_0}{2}\right)^2 + \frac{dy_0^2}{4},$$

и требуемое представление возможно только при чётном y_0.

При $d = 3$ заменой δ на $\omega\delta$ или на $\omega^2\delta$ можно сделать y_0 чётным. При $d = 7$ число y_0 обязано быть чётным. А вот при $d = 11$ число y_0 уже может оказаться нечётным, и тогда искомое представление (1.1) будет невозможным. Например, для $p = 23$ имеем

$$23 = 4^2 + 4 \cdot 1 + 3 \cdot 1^2 = 5^2 + 5 \cdot (-1) + 3 \cdot (-1)^2,$$

так что $y_0 = \pm 1$ — нечётное число.

Пример 1.2. Пусть $d = 1$, $p = 2017$. Тогда можно взять $m_0 = 229$. Положим $\rho_0 = 2017$, $\rho_1 = 229 + \sqrt{-1}$ и вычислим $\delta = \text{НОД}\,(\rho_0, \rho_1)$. Имеем

$$\frac{\rho_0}{\rho_1} = \frac{229 - \sqrt{-1}}{26} \approx 8{,}80 - 0{,}03\sqrt{-1}, \quad \gamma^* = 9,$$

так что $\rho_2 = \rho_0 - \rho_1\gamma^* = -44 - 9\sqrt{-1}$. Далее находим

$$\frac{\rho_1}{\rho_2} = -5 + \sqrt{-1} = \gamma^*$$

и $\rho_3 = \rho_1 - \rho_2\gamma^* = 0$. Значит, $\delta = \rho_2 = -44 - 9\sqrt{-1}$ и $2017 = 9^2 + 44^2$.

1.3. Сумма символов Лежандра

Пусть b и c — целые числа. Следующее утверждение о значении суммы

$$S(b,c)=\sum_{x=0}^{p-1}\Big(\frac{x^2+bx+c}{p}\Big)$$

составляет основу всех дальнейших вычислений.

Лемма 1.5. *Справедливо равенство*

$$S(b,c)=\begin{cases} p-1, & \text{если } D\equiv 0 \pmod p,\\ -1, & \text{если } D\not\equiv 0 \pmod p,\end{cases} \tag{1.4}$$

где $D=b^2-4c$.

Доказательство. Выделяя в выражении x^2+bx+c полный квадрат, нетрудно обнаружить, что

$$S(b,c)=S(0,-D)=\sum_{x=0}^{p-1}\Big(\frac{x^2-D}{p}\Big).$$

Сначала вычислим $S(0,-D)$ по модулю p, воспользовавшись критерием Эйлера. Имеем

$$S(0,-D)\equiv\sum_{x=0}^{p-1}(x^2-D)^{(p-1)/2}=\sum_{s=0}^{p^*}C_{p^*}^s(-D)^{p^*-s}\sum_{x=0}^{p-1}x^{2s}\pmod p,$$

где $p^*=(p-1)/2$. Далее нам понадобится следующая формула для степенной суммы по модулю p (здесь t — целое число):

$$\sum_{x=1}^{p-1}x^t\equiv_p\begin{cases}-1, & \text{если } t \text{ делится на } p-1,\\ 0 & \text{иначе}\end{cases}$$

(несложное доказательство, основанное на цикличности *мультипликативной группы* поля $\mathbb{Z}_p$, предоставляется читателю). Тогда

$$\sum_{x=0}^{p-1}x^{2s}\equiv_p\begin{cases}0 & \text{при } 0\leqslant s<p^*,\\ -1 & \text{при } s=p^*.\end{cases}$$

Следовательно, $S(0,-D)\equiv -1 \pmod p$. Теперь, поскольку $|S(0,-D)|\leqslant p$, получим $S(0,-D)\in\{-1,p-1\}$ для любого D. Ясно, что $S(0,0)=p-1$. Но

$$\sum_{D=0}^{p-1}S(0,-D)=\sum_{x=0}^{p-1}\sum_{D=0}^{p-1}\Big(\frac{x^2-D}{p}\Big)=0,$$

поэтому $S(0,-1)=S(0,-2)=\ldots=S(0,-p+1)=-1$. □

Существуют и другие способы доказательства формулы (1.4) (в нашем доказательстве мы следовали Якобсталю [21]). Вычисление суммы $S(0,-D)$ эквивалентно подсчёту числа точек гиперболы $y^2 = x^2 - D$ над полем $\mathbb{Z}_p$. С точки зрения элементарной алгебраической геометрии, это кривая второго порядка, которая допускает *рациональную параметризацию*. В данном случае удобно перейти к новым переменным $u = x - y$, $v = x + y$ и записать уравнение в виде

$$uv = D.$$

Линейная замена $(x, y) \to (u, v)$ биективна, а число решений (u, v) последнего уравнения над полем $\mathbb{Z}_p$ легко находится в зависимости от D. Таким образом, имеем

$$\sum_{x=0}^{p-1}\Big(1 + \Big(\frac{x^2-D}{p}\Big)\Big) = \begin{cases} 2p-1, & \text{если } D \equiv 0 \pmod p, \\ p-1, & \text{если } D \not\equiv 0 \pmod p. \end{cases}$$

Отсюда и следует равенство (1.4).

§ 2. Вычисление суммы $\phi(n)$

При замене x на $-x$ выражение $x^3 + nx$ меняет знак, что позволяет записать сумму $\phi(n)$ в виде

$$\phi(n) = \Big(1 + \Big(\frac{-1}{p}\Big)\Big) \sum_{x=1}^{(p-1)/2} \Big(\frac{x^3+nx}{p}\Big).$$

В случае $p \equiv 3 \pmod 4$ имеем $\Big(-\frac{1}{p}\Big) = -1$, поэтому $\phi(n) = 0$ для любого n. В частности, для эллиптической кривой E, заданной уравнением

$$y^2 = x^3 + nx, \tag{2.1}$$

имеем $N_p(E) = p + 1$.

Далее в этом разделе будем рассматривать только случай $p \equiv 1 \pmod 4$. В этом случае $\Big(-\frac{1}{p}\Big) = 1$ и $\phi(n)$ чётно при любом n. Ясно также, что $\phi(n) = 0$ при $n \equiv 0 \pmod p$.

Введём обозначение: $k = (p-1)/4$, так что $p = 4k + 1$ и $(p-1)/2 = 2k$.

2.1. Абсолютное значение $\phi(n)$

Пусть $m \not\equiv 0 \pmod p$. Имеем

$$\phi(nm^2) = \sum_{x=0}^{p-1}\Big(\frac{x^3+nm^2x}{p}\Big) = \Big(\frac{m}{p}\Big)\sum_{x=0}^{p-1}\Big(\frac{x^3+nx}{p}\Big) = \Big(\frac{m}{p}\Big)\phi(n).$$

Как следствие, получим

$$|\phi(nm^2)| = |\phi(n)|.$$

Отсюда видно, что $|\phi(n)|$ при $n \not\equiv 0 \pmod p$ может принимать только два значения: одно для квадратичных вычетов n по модулю p, другое — для квадратичных невычетов. Более конкретно можно выразиться так: либо $|\phi(n)| = |\phi(1)|$, либо $|\phi(n)| = |\phi(g)|$, где g — какой-нибудь *первообразный корень* по модулю p.

Рассмотрим частный случай $m = m_0$, где $m_0^2 \equiv -1 \pmod p$. Имеем

$$\phi(-n) = \phi(nm_0^2) = \left(\frac{m_0}{p}\right)\phi(n) = (-1)^k\phi(n),$$

поскольку по критерию Эйлера

$$\left(\frac{m_0}{p}\right) \equiv m_0^{2k} \equiv (-1)^k \pmod p.$$

В частности, $\phi(1) = (-1)^k\phi(-1)$ (это равенство нам понадобится в подразделе 2.3).

2.2. Значение $\phi(n)$ по модулю p

Важным шагом на пути вычисления $\phi(n)$ является вычисление $\phi(n) \bmod p$. Применим ту же технику, что и при доказательстве леммы (1.5):

$$\phi(n) \equiv \sum_{x=0}^{p-1}(x^3+nx)^{(p-1)/2} = \sum_{x=0}^{p-1}\sum_{s=0}^{2k} C_{2k}^s x^{2k+2s} n^{2k-s} = \\ = \sum_{s=0}^{2k} C_{2k}^s n^{2k-s} \sum_{x=0}^{p-1} x^{2k+2s} \equiv -n^k C_{2k}^k \pmod p.$$

Как следствие, получим

$$\frac{\phi(n)}{2} \equiv n^k a \pmod p, \tag{2.2}$$

где введено обозначение $a = \phi(1)/2$. Формула (2.2) в виде формулы (10) есть в статье [24] (см. также статью [30], где по модулю p найдены суммы Якобсталя (0.2) для любого l).

Главное наблюдение: если удастся вычислить a, то с помощью сравнения (2.2) можно однозначно и быстро найти $\phi(n)/2$ для любого n, поскольку a priori имеет место оценка

$$\left|\frac{\phi(n)}{2}\right| \leqslant \frac{p-1}{2}$$

и $n^k \bmod p$ быстро вычисляется с помощью бинарного алгоритма.

Таким образом, всё сводится к нахождению числа a. Как это можно сделать, довольно понятно написано ещё самим Якобсталем в статье [21]. Следующие подразделы 2.3 и 2.4 представляют собой пересказ соответствующей части этой статьи.

2.3. Значение $\phi(1)/2$ по модулю 4

Якобсталь фактически вычислил значение $a = \phi(1)/2$ по модулю 8 (см. ниже (2.5)). Удобно сначала рассмотреть $a' = \phi(-1)/2$. Имеем

$$2a' = \sum_{x=0}^{p-1}\Big(\frac{x^3-x}{p}\Big) = \sum_{x=1}^{p-3}\Big(\frac{x}{p}\Big)\Big(\frac{x+1}{p}\Big)\Big(\frac{x+2}{p}\Big). \tag{2.3}$$

Пусть $N_p^{(3)}$ — число тех $x \in \{1, 2, \ldots, p-3\}$, для которых

$$\Big(\frac{x}{p}\Big) = \Big(\frac{x+1}{p}\Big) = \Big(\frac{x+2}{p}\Big) = -1.$$

Тогда

$$8N_p^{(3)} = \sum_{x=1}^{p-3}\Big(1-\Big(\frac{x}{p}\Big)\Big)\Big(1-\Big(\frac{x+1}{p}\Big)\Big)\Big(1-\Big(\frac{x+2}{p}\Big)\Big). \tag{2.4}$$

Кроме того, число $N_p^{(3)}$ чётно (поскольку из $x \in N_p^{(3)}$ следует $p-x-2 \in N_p^{(3)}$, так что все элементы $N_p^{(3)}$ разбиваются на пары). Теперь из равенств (2.3) и (2.4) с помощью леммы 1.5 можно вывести равенство $8N_p^{(3)} = p-3-2a'$ (это тривиальное, но несколько громоздкое и скучное упражнение решается исключительно усилием воли; вслед за Якобсталем мы предоставим это читателю).

Как следствие, приходим к сравнению

$$a' \equiv \frac{p-3}{2} = 2k-1 \pmod 8.$$

Далее имеем

$$a = (-1)^k a' \equiv (-1)^k(2k-1) \pmod 8, \tag{2.5}$$

откуда $a \equiv -1 \pmod 4$, поскольку $(-1)^k(2k-1) \equiv -1 \pmod 4$ при любом k.

Таким образом, если нам удастся вычислить $|a|$, то с помощью полученного сравнения можно однозначно определить и само число a.

2.4. Связь с представлением $p = A^2 + B^2$

Последнее вычисление, которые мы предпримем, является, пожалуй, самым важным: оно приводит к формуле (2.6), лежащей в основе любого эффективного алгоритма вычисления $\phi(n)$.

Читателю предлагается убедиться в справедливости следующей цепочки равенств:

$$\begin{aligned}\sum_{n=0}^{p-1}\phi^2(n) &= \sum_{n=0}^{p-1}\sum_{x=1}^{p-1}\sum_{y=1}^{p-1}\Big(\frac{x^3+nx}{p}\Big)\Big(\frac{y^3+ny}{p}\Big) = \\ &= \sum_{x=1}^{p-1}\sum_{y=1}^{p-1}\Big(\frac{xy}{p}\Big)\sum_{n=0}^{p-1}\Big(\frac{n^2+(x^2+y^2)n+x^2y^2}{p}\Big) = \\ &= \sum_{x=1}^{p-1}\sum_{y=1}^{p-1}\Big(\frac{xy}{p}\Big)\sum_{n=0}^{p-1}\Big(\frac{n^2-(x^2-y^2)^2}{p}\Big) = \\ &= p\sum_{x=1}^{p-1}\Big(\frac{x^2}{p}\Big)+p\sum_{x=1}^{p-1}\Big(\frac{-x^2}{p}\Big)-\sum_{x=1}^{p-1}\sum_{y=1}^{p-1}\Big(\frac{xy}{p}\Big) = 2p(p-1).\end{aligned}$$

Теперь, учитывая результат подраздела 2.1, получим

$$\sum_{n=0}^{p-1}\phi^2(n) = \frac{p-1}{2}\big(\phi^2(1)+\phi^2(g)\big) = 2p(p-1),$$

где g — первообразный корень по модулю p. Отсюда

$$p = \Big(\frac{\phi(1)}{2}\Big)^2+\Big(\frac{\phi(g)}{2}\Big)^2. \tag{2.6}$$

В частности, $|\phi(1)| < 2p^{1/2}$ и $|\phi(g)| < 2p^{1/2}$, так что $|\phi(n)| < 2p^{1/2}$ вообще для любого $n \not\equiv 0 \pmod p$. Тем самым полностью доказано утверждение теоремы Хассе для эллиптической кривой E, заданной уравнением (2.1).

Равенство (2.6) есть не что иное, как представление числа p в виде суммы двух квадратов. Таким образом, нам осталось найти альтернативный (и эффективный на практике) способ получения такого представления. Как мы уже видели (см. подразделы 1.1 и 1.2), такой способ есть, и даже не один.

2.5. Быстрый алгоритм вычисления $\phi(n)$

Представленные выше факты приводят к следующему алгоритму вычисления $\phi(n)$ при $n \not\equiv 0 \pmod p$, работающему в случае больших p.

(а) С помощью какого-нибудь быстрого алгоритма (см. подразделы 1.1 и 1.2) находим представление простого $p = 4k+1$ в виде суммы двух квадратов:

$$p = A^2+B^2, \tag{2.7}$$

где A, B — натуральные числа, при этом A нечётно. Как видно из леммы 1.1, такие числа A и B определяются числом p однозначно.

(б) Затем находим $a = \phi(1)/2$, исходя из условий

$$|a| = A, \quad a \equiv -1 \pmod 4.$$

(в) Наконец, для заданного n находим $\phi(n)$, исходя из условий

$$\frac{\phi(n)}{2} \equiv n^k a \pmod p, \quad \left|\frac{\phi(n)}{2}\right| \leqslant \frac{p-1}{2}.$$

Проиллюстрируем данный алгоритм одним примером.

Пример 2.1. Пусть $p{=}2017{=}4{\cdot}504{+}1$ и $n{=}-37$. Имеем $2017{=}9^2{+}44^2$ (см. пример 1.2), так что $A = 9$ и, таким образом, $a = -9$. Тогда

$$\frac{\phi(-37)}{2} \equiv (-37)^{504} \cdot (-9) \equiv 1973 \pmod{2017}.$$

Поскольку $1973 > 1008 = (2017-1)/2$, получим

$$\frac{\phi(-37)}{2} = 1973 - 2017 = -44,$$

откуда $\phi(-37) = -88$.

Упражнение 2.1. Для простого числа $p = 4k+1$ опишите алгоритм быстрого вычисления биномиального коэффициента C_{2k}^k по модулю p.

Указание. Воспользуйтесь сравнением $C_{2k}^k \equiv -\phi(1) \pmod p$.

Упражнение 2.2. Докажите, что для простого числа $p = 4k+1$ представление (2.7) может быть найдено с помощью следующих *формул Гаусса*:

$$A \equiv \frac{1}{2} C_{2k}^k \pmod p, \quad B \equiv (2k)!A \pmod p,$$

где (не обязательно положительные) числа A, B удовлетворяют дополнительным условиям $|A| \leqslant (p-1)/2$, $|B| \leqslant (p-1)/2$.

Комментарий. Формулы Гаусса непригодны для непосредственного разложения больших чисел p в сумму двух квадратов, поскольку непонятно, как быстро вычислить C_{2k}^k по модулю p, не прибегая к алгоритму из упражнения 2.1.

Упражнение 2.3. Пусть $N_p(G)$ — число точек эллиптической кривой G, заданной уравнением $u^2v^2 + u^2 + v^2 = 1$, над полем $\mathbb{Z}_p$. Докажите, что

$$N_p(G) = N_p(E) - 2\Big(\frac{-1}{p}\Big) = p + 1 + \phi(4) - 2\Big(\frac{-1}{p}\Big), \tag{2.8}$$

где E — кривая (2.1) с $n = 4$.

Указание. Воспользуйтесь *бирациональной заменой*

$$u=\frac{2x}{y},\quad v=\frac{x-2}{x+2}\quad\leftrightarrows\quad x=\frac{2(1+v)}{1-v},\quad y=\frac{4(1+v)}{u(1-v)}.$$

Комментарий. В формуле (2.8) учитываются и две точки кривой G на бесконечности (это точки пересечения G с бесконечно удалённой прямой). Формула для $N_p(G)$ впервые встречается в дневниках Гаусса как предположение, которое позднее было доказано (см. по этому поводу [15], а также комментарий на стр. 97 в книге [11]).

Для сравнения укажем ещё один алгоритм для вычисления $\phi(n)$, который можно предложить на основе теоремы 6.2.1 из книги [13]. Он также опирается на представление (2.7), однако теперь в вычислениях будут участвовать как A, так и B.

(а) Выберем какой-нибудь первообразный корень g по модулю p.

(б) Определим a_4, исходя из условий

$$a_4\equiv-\Big(\frac{2}{p}\Big)\pmod 4,\quad |a_4|=A.$$

(в) Затем вычислим b_4, исходя из условий

$$b_4\equiv g^k a_4\pmod p,\quad |b_4|=B.$$

(г) Далее найдём l, для которого $g^l\equiv n\pmod p$, и вычислим $r=l \bmod 4$.

(д) Тогда

$$\phi(n)=\begin{cases}2(-1)^k a_4, & \text{если } r=0,\\ 2(-1)^k b_4, & \text{если } r=1,\\ 2(-1)^{k+1} a_4, & \text{если } r=2,\\ 2(-1)^{k+1} b_4, & \text{если } r=3.\end{cases}\tag{2.9}$$

Для доказательства корректности этого алгоритма достаточно проверить, что для числа $\phi(n)$, найденного по формуле (2.9), выполняется сравнение (2.2). Проще всего рассмотреть все ситуации в зависимости от значений $(2/p)$ и r. Пусть, например, мы имеем

$$\Big(\frac{2}{p}\Big)=1,\quad r=3.$$

В этом случае $a_4=a$, а число k чётно. Тогда по формуле (2.9) получим

$$\frac{\phi(n)}{2}=-b_4\equiv-g^k a\pmod p.$$

Нам нужно убедиться, что $-g^k a\equiv n^k a\pmod p$. Действительно, имеем

$$n^k\equiv g^{kl}=g^{k(4m+3)}\equiv g^{3k}\equiv-g^k\pmod p,$$

поскольку $g^{2k}=g^{(p-1)/2}\equiv-1\pmod p$.

По очевидным причинам (сначала требуется найти g, а затем решать вычислительно сложную задачу *дискретного логарифмирования*, чтобы найти l и r) данный алгоритм не будет эффективным при больших p.

§ 3. Вычисление суммы $\psi(n)$

В этом разделе мы рассмотрим сумму (0.3), вычисление которой можно организовать по тому же сценарию, что и вычисление суммы (0.1). Поэтому мы подробно опишем только наиболее сложные этапы вычисления, предоставив читателю самому восстановить детали в остальных случаях.

I. Сначала заметим, что при $p \equiv 2 \pmod 3$ отображение

$$f : \mathbb{Z}_p \to \mathbb{Z}_p, \quad f(x) = x^3,$$

биективно (очередное упражнение для читателя), поэтому

$$\psi(n) = \sum_{x=0}^{p-1} \Big(\frac{x^3+n}{p}\Big) = \sum_{y=0}^{p-1} \Big(\frac{y+n}{p}\Big) = 0.$$

Далее пусть $p \equiv 1 \pmod 3$. Положим $k = (p-1)/6$, так что $p = 6k+1$ и $(p-1)/2 = 3k$. Пусть также $n \not\equiv 0 \pmod p$ (иначе, очевидно, $\psi(n) = 0$).

Так же как и в подразделе 2.1 легко доказать, что

$$\psi(nm^3) = \Big(\frac{m}{p}\Big)\psi(n) \tag{3.1}$$

при любом $m \not\equiv 0 \pmod p$. Отсюда следует, что $|\psi(n)|$ принимает одно из трёх значений $|\psi(g^u)|$, $|\psi(g^v)|$ и $|\psi(g^w)|$, где u, v и w — какая-нибудь полная система вычетов по модулю 3 (как и выше, g — фиксированный первообразный корень по модулю p). Далее мы возьмём $u = 0$, $v = 2$, $w = 4$.

II. Теперь заметим, что число $\psi(n)$ чётно тогда и только тогда, когда

$$n^k \equiv \pm 1 \pmod p. \tag{3.2}$$

Действительно, при выполнении условия (3.2) сравнение

$$x^3 + n \equiv 0 \pmod p$$

имеет ровно три решения, а иначе оно неразрешимо. Значит, среди слагаемых $\Big(\frac{x^3+n}{p}\Big)$ суммы $\psi(n)$ число плюс-минус единиц равно либо $p-3$, либо p, что и доказывает утверждение. В частности, число $\psi(1)$ чётно, а числа $\psi(g^2)$ и $\psi(g^4)$ нечётны.

III. Следующий шаг — *вычисление* $\psi(n)$ по модулю p. Имеем

$$\psi(n) \equiv \sum_{x=0}^{p-1}(x^3+n)^{(p-1)/2} = \sum_{x=0}^{p-1}\sum_{s=0}^{3k} C_{3k}^s x^{3s} n^{3k-s} = \\ = \sum_{s=0}^{3k} C_{3k}^s n^{3k-s} \sum_{x=0}^{p-1} x^{3s} \equiv -n^k C_{3k}^{2k} \pmod p.$$

В частности, видно, что всегда $\psi(n) \not\equiv 0 \pmod p$. Кроме того,

$$\psi(n) \equiv 2n^k a \pmod p, \tag{3.3}$$

где используется обозначение $a = \psi(1)/2$. Если удастся вычислить число a, то, учитывая чётность числа $\psi(n)$, с помощью сравнения (3.3) и неравенства $|\psi(n)| \leqslant p$ мы сможем однозначно определить $\psi(n)$.

IV. Как мы уже знаем, число a — целое. Более того, справедливо сравнение

$$a \equiv -1 \pmod 3. \tag{3.4}$$

В самом деле, имеем

$$\psi(n) = \Big(\frac{n}{p}\Big) + \sum_{x=1}^{p-1}\Big(\frac{x^3+n}{p}\Big) = \Big(\frac{n}{p}\Big) + \sum_{y=1}^{p-1} N_p(y)\Big(\frac{y+n}{p}\Big),$$

где $N_p(y)$ — число решений сравнения $x^3 \equiv y \pmod p$. При $y \not\equiv 0 \pmod p$ имеем $N_p(y) \in \{0, 3\}$, поэтому

$$\psi(n) \equiv \Big(\frac{n}{p}\Big) \pmod 3.$$

В частности, при $n = 1$ отсюда следует сравнение (3.4).

V. Осталось самое сложное — найти связь числа a с представлением

$$p = A^2 + 3B^2, \tag{3.5}$$

где натуральные A и B однозначно определены числом p (лемма 1.1). Мы докажем, что $|a| = A$ и, следовательно, $a = \pm A$, где знак выбирается в соответствии с (3.4).

С этой целью вычислим две суммы:

$$S_1 = \sum_{n=0}^{p-1} \psi(n^2), \quad S_2 = \sum_{n=0}^{p-1} \psi^2(n).$$

Основным инструментом при вычислении будет лемма 1.5.

Первая сумма вычисляется так:

$$S_1 = \sum_{n=0}^{p-1}\sum_{x=0}^{p-1}\Big(\frac{x^3+n^2}{p}\Big) = \sum_{x=0}^{p-1}\sum_{n=0}^{p-1}\Big(\frac{n^2+x^3}{p}\Big) = p-1+(-1)(p-1) = 0.$$

С другой стороны, имеем

$$S_1 = 2\sum_{j=0}^{3k-1}\psi(g^{2j}) = 2\sum_{l=0}^{k-1}\big(\psi(g^{6l})+\psi(g^{6l+2})+\psi(g^{6l+4})\big) = \\ = 2k\big(\psi(1)+\psi(g^2)+\psi(g^4)\big),$$

поскольку

$$\psi(g^{6l}) = \psi(1), \quad \psi(g^{6l+2}) = \psi(g^2), \quad \psi(g^{6l+4}) = \psi(g^4)$$

(см. формулу (3.1)). В качестве следствия получим равенство

$$\psi(1)+\psi(g^2)+\psi(g^4) = 0. \tag{3.6}$$

Вычисление второй суммы можно организовать следующим образом:

$$S_2 = \sum_{n=0}^{p-1}\sum_{x=0}^{p-1}\sum_{y=0}^{p-1}\Big(\frac{x^3+n}{p}\Big)\Big(\frac{y^3+n}{p}\Big) = \\ = \sum_{x=0}^{p-1}\sum_{y=0}^{p-1}\sum_{n=0}^{p-1}\Big(\frac{n^2+(x^3+y^3)n+x^3y^3}{p}\Big) = \\ = \sum_{x=0}^{p-1}\sum_{y=0}^{p-1}\sum_{n=0}^{p-1}\Big(\frac{n^2-(x^3-y^3)^2}{p}\Big) = (p-1)N_p+(-1)(p^2-N_p),$$

где N_p — число решений (x, y) сравнения $x^3 - y^3 \equiv 0 \pmod p$. Нетрудно видеть, что

$$N_p = 3(p-1)+1 = 3p-2,$$

поэтому окончательно получим

$$S_2 = 2p(p-1).$$

С другой стороны, имеем

$$S_2 = \sum_{n=0}^{p-1}\psi^2(n) = 2k\big(\psi^2(1)+\psi^2(g^2)+\psi^2(g^4)\big).$$

Следовательно,

$$\psi^2(1)+\psi^2(g^2)+\psi^2(g^4) = 6p. \tag{3.7}$$

Наконец, из полученных равенств (3.6) и (3.7) следует равенство

$$\psi^2(1) + \psi(1)\psi(g^2) + \psi^2(g^2) = 3p,$$

которое можно переписать в виде

$$p = \Big(\frac{\psi(1)}{2}\Big)^2 + 3\Big(\frac{\psi(1) + 2\psi(g^2)}{6}\Big)^2.$$

Таким образом, $|a| = A$ в силу единственности представления (3.5).

VI. На основе сказанного выше можно предложить следующий алгоритм вычисления $\psi(n)$ при любом $n \not\equiv 0 \pmod p$.

(а) С помощью одного из алгоритмов (см. подразделы 1.1 и 1.2) находим представление простого $p = 6k + 1$ в виде (3.5).

(б) Находим $a = \psi(1)/2$, исходя из равенства $|a| = A$ и сравнения (3.4).

(в) С помощью сравнения (3.2) определяем чётность $\psi(n)$, а затем находим $\psi(n)$, опираясь на сравнение (3.3) и неравенство $|\psi(n)| \leqslant p - 1$.

Пример 3.1. Пусть $p = 2017 = 6 \cdot 336 + 1$ и $n = -432$. Имеем

$$2017 = 17^2 + 3 \cdot 24^2$$

(см. пример 1.1), так что $A = 17$ и, следовательно, $a = 17$. Поскольку

$$(-432)^{336} \equiv 1 \pmod{2017},$$

число $\psi(-432)$ чётно, при этом

$$\psi(-432) \equiv 2 \cdot (-432)^{336} \cdot 17 \equiv 34 \pmod{2017}.$$

Значит, $\psi(-432) = 34$.

Упражнение 3.1. Докажите утверждение теоремы Хассе для эллиптической кривой E, заданной уравнением

$$y^2 = x^3 + n. \tag{3.8}$$

Упражнение 3.2. Пусть $p > 3$ и $N_p(F)$ — число точек (эллиптической) *кривой Ферма* F, заданной уравнением $u^3 + v^3 = 1$, над полем $\mathbb{Z}_p$. Докажите формулу

$$N_p(F) = N_p(E) = p + 1 + \psi(-432), \tag{3.9}$$

где E — кривая (3.8) с $n = -432$.

Указание. Воспользуйтесь бирациональной заменой

$$u = \frac{36 + y}{6x}, \quad v = \frac{36 - y}{6x} \quad \leftrightarrows \quad x = \frac{12}{u + v}, \quad y = \frac{36(u - v)}{u + v}.$$

Комментарий. Формула (3.9) учитывает точки (одну или три) кривой F на бесконечности. При $p \equiv 1 \pmod 3$ имеем

$$\psi(-432) = \psi((-3)^3 \cdot 16) = \left(\frac{-3}{p}\right)\psi(16) = \psi(16),$$

а если $p \equiv 2 \pmod 3$, то $\psi(-432) = 0$.

Упражнение 3.3. Докажите *теорему Гаусса* (см. § 8.3 в книге [1], а также § 2 главы IV в книге [29]):
если $p \equiv 1 \pmod 3$ и в представлении

$$4p = C^2 + 27D^2 \tag{3.10}$$

имеем $C \equiv 1 \pmod 3$, то

$$N_p(F) = p + 1 + C.$$

Попутно выясните, когда число 2 будет *кубическим вычетом* по модулю p.

Решение. Прежде всего заметим, что представление (3.10) возможно, при этом числа C и D определены однозначно с точностью до знака (это можно вывести из аналогичного утверждения о представлении (3.5)). Как следствие, сравнение $C \equiv 1 \pmod 3$ определяет число C однозначно.

I. Пусть $N_p = N_p(F) - 3 = p - 2 + \psi(16)$ — число решений сравнения

$$u^3 + v^3 \equiv 1 \pmod p.$$

Тогда, как нетрудно обнаружить, $N_p \equiv 6 \pmod 9$. Следовательно, верно сравнение

$$4p + 4\psi(16) \equiv 5 \pmod 9.$$

Поскольку

$$\psi(16) \equiv \psi(1) \equiv \psi(g^2) \equiv \psi(g^4) \equiv 1 \pmod 3,$$

возможны три случая: либо $\psi(16) = \psi(1)$, либо $\psi(16) = \psi(g^2)$, либо $\psi(16) = \psi(g^4)$.

1) Пусть $\psi(16) = \psi(1)$. Имеем

$$4p + 4\psi(16) = \psi^2(1) + 3\left(\frac{\psi(1) + 2\psi(g^2)}{3}\right)^2 + 4\psi(1) \equiv 5 \pmod 9.$$

Так как $\psi(1) \equiv 1 \pmod 3$, имеем $\psi^2(1) + 4\psi(1) \equiv 5 \pmod 9$. Значит,

$$\frac{\psi(1) + 2\psi(g^2)}{3} \equiv 0 \pmod 3$$

и, таким образом, $C = \psi(1)$.

2) Пусть теперь $\psi(16) = \psi(g^2)$. В этом случае имеем

$$4p + 4\psi(16) = \psi^2(g^2) + 3\left(\frac{2\psi(1) + \psi(g^2)}{3}\right)^2 + 4\psi(g^2) \equiv 5 \pmod 9.$$

Как и выше, отсюда следует, что

$$\frac{2\psi(1) + \psi(g^2)}{3} \equiv 0 \pmod 3$$

и, таким образом, $C = \psi(g^2)$.

3) В случае $\psi(16) = \psi(g^4)$ рассуждения аналогичны и мы получим $C = \psi(g^4)$.

II. Очевидно, число 2 является кубическим вычетом по модулю p тогда и только тогда, когда $\psi(16) = \psi(1)$. Поскольку число $\psi(1)$ чётно, а числа $\psi(g^2)$ и $\psi(g^4)$ нечётны, последнее условие выполнено тогда и только тогда, когда $C \equiv 0 \pmod 2$ в представлении (3.10) (или, в терминах представления (3.5), когда $B \equiv 0 \pmod 3$).

§ 4. Заключительные замечания

Важно подчеркнуть, что рассмотренные нами быстрые алгоритмы вычисления сумм Якобсталя $\phi(n)$ и $\psi(n)$ имеют вероятностную природу, поскольку их ключевой ингредиент — некоторое решение $x = m_0$ сравнения (1.2) — предполагается находить вероятностными методами. В связи с этим возникает вопрос: существуют ли быстрые *детерминированные* алгоритмы извлечения квадратных корней по модулю простого числа p? Ответ, как выясняется, положительный, но неожиданный: для решения этой задачи можно приспособить уже упомянутый нами алгоритм Шуфа, который находит число точек эллиптической кривой над конечным полем и имеет *полиномиальную* сложность (см. оригинальную работу [26]). При этом даже не приходится опираться на правдоподобные, но ещё не доказанные гипотезы в теории чисел, как это иногда бывает.

Так, в предположении верности *обобщённой гипотезы Римана* квадратичный невычет b для алгоритма Тонелли — Шенкса можно найти простым перебором за полиномиальное время. Как следствие, алгоритм[7)] Тонелли — Шенкса становится детерминированным и полиномиальным.

Таким образом, если нас интересует безусловный (не апеллирующий ни к каким гипотезам), детерминированный и полиномиальный алгоритм вычисления классических сумм Якобсталя, то на данный момент

[7)] К слову, этот алгоритм правильнее было бы называть алгоритмом Тонелли, ибо Шенкс лишь переоткрыл его спустя 80 лет, не наведя исторические справки по анекдотичной причине (см. [36]).

можно предложить только алгоритм Шуфа. Который, однако, довольно сложен теоретически и, как уже отмечалось, на практике проигрывает более наивным вероятностным алгоритмам.

Классические суммы Якобсталя отвечают за число точек на очень специальных эллиптических кривых (0.4) над полем $\mathbb{Z}_p$. Про такие кривые говорят, что они допускают *комплексное умножение* (что это такое, можно узнать только основательно погрузившись в теорию эллиптических кривых). Естественно, что в *системах компьютерной алгебры* (например PARI/GP) для подсчёта точек на эллиптических кривых с комплексным умножением над конечными полями используются быстрые практичные алгоритмы типа тех, что были рассмотрены в настоящей статье[8)]. И только в более сложных случаях применяется алгоритм Шуфа и его модификации (подробности можно узнать по ссылке [37]).

Фундаментальное изучение общих сумм Якобсталя $\phi_l(n)$ (включая и рассмотренные нами случаи первых значений l) возможно на основе более сложных конструкций — сумм Гаусса и Якоби. Систематическое изложение результатов в этой области читатель может найти в монографии [13]. Для первоначального знакомства с указанными суммами могут быть полезны соответствующие разделы в книгах [1] и [9].

Автор выражает благодарность А. В. Устинову за содержательные замечания и комментарии по теме статьи.

Список литературы

[1] *Айерлэнд К.*, *Роузен М.* Классическое введение в современную теорию чисел. М.: Мир, 1987.

[2] *Василенко О. Н.* Теоретико-числовые алгоритмы в криптографии. М.: МЦНМО, 2006.

[3] *Винберг Э. Б.* Курс алгебры. М.: МЦНМО, 2019.

[4] *Вялый М. Н.* О представлении чисел в виде суммы двух квадратов // Математическое просвещение. Сер. 3. Вып. 10. М.: МЦНМО. 2006. С. 190–194.

[5] *Дэвенпорт Г.* Высшая арифметика. М.: Наука, 1965.

[6] *Кнэпп Э.* Эллиптические кривые. М.: Факториал Пресс, 2004.

[7] *Коблиц Н.* Курс теории чисел и криптографии. М.: ТВП, 2001.

[8] *Кострикин А. И.* Введение в алгебру. Часть 3. Основные структуры. М.: МЦНМО, 2018.

[9] *Лидл Р.*, *Нидеррайтер Г.* Конечные поля. М.: Мир, 1988.

[8)] На самом деле всё не так просто и совсем не элементарно (читатель может заглянуть в статью [25]).

[10] *Нестеренко Ю. В.* Теория чисел. М.: Академия, 2008.

[11] *Степанов С. А.* Арифметика алгебраических кривых. М.: Наука, 1991.

[12] *Basilla J. M.* On the solution of $x^2 + dy^2 = m$ // Proc. Japan Acad. Ser. A Math. Sci 2004. Vol. 80, № 5. P. 40–41.

[13] *Berndt B. C.*, *Evans R. J.*, *Williams K. S.* Gauss and Jacobi sums. New York: John Wiley & Sons, Inc., 1998.

[14] *Brillhart J.* Note on representing a prime as a sum of two squares // Math. Comp. 1972. Vol. 26. P. 1011–1013.

[15] *Chowla S.* The last entry in Gauss's diary // Proc. Nat. Acad. Sci. U.S.A. 1949. Vol. 35, № 5. P. 244–246.

[16] *Clarke F. W.*, *Everitt W. N.*, *Littlejohn L. L.*, *Vorster S. J. R.* H. J. S. Smith and the Fermat two squares theorem // Amer. Math. Monthly. 1999. Vol. 106, № 7. P. 652–665.

[17] *Cohen H.* A course in computational algebraic number theory. Berlin: Springer-Verlag, 1993. (Grad. Texts Math.; Vol. 138).

[18] *Cox D. A.* Primes of the form $x^2 + ny^2$. New York: John Wiley & Sons, Inc., 1989.

[19] *Evink T.*, *Helminck P. A.* Tribonacci numbers and primes of the form $p = x^2 + 11y^2$ // https://arxiv.org/abs/1801.04605

[20] *Hermite C.* Note au sujet de l'article précédent // J. Math. Pures Appl. 1848. Vol. 13. P. 15.

[21] *Jacobsthal E.* Über die Darstellung der Primzahlen der Form $4n + 1$ als Summe zweier Quadrate // J. Reine Angew. Math. 1907. Bd. 132. S. 238–246.

[22] *Katre S. A.* Jacobsthal sums in terms of quadratic partitions of a prime // Number theory (Ootacamund, 1984). Berlin: Springer-Verlag, 1985. (Lect. Notes Math.; Vol. 1122). P. 153–162.

[23] *Koblitz N.* Algebraic aspects of cryptography. Berlin: Springer-Verlag, 1998. (Alg. Comp. Math.; Vol. 3).

[24] *Lehmer E.* On Euler's criterion // J. Austral. Math. Soc. 1959/1961. Vol. 1, part 1. P. 64–70.

[25] *Rubin K.*, *Silverberg A.* Point counting on reductions of CM elliptic curves // J. Number Theory. 2009. Vol. 129, № 12. P. 2903–2923.

[26] *Schoof R.* Elliptic curves over finite fields and the computation of square roots mod p // Math. Comp. 1985. Vol. 44, №. 170. P. 483–494.

[27] *von Schrutka L.* Ein Beweis für die Zerlegbarkeit der Primzahlen von der Form $6n + 1$ in ein einfaches und ein dreifaches Quadrat // J. Reine Angew. Math. 1911. Bd. 140. S. 252–265.

[28] *Serret J.-A.* Sur un théorème relatif aux nombres entieres // J. Math. Pures Appl. 1848. Vol. 13. P. 12–14.

[29] *Silverman J. H.*, *Tate J.* Rational points on elliptic curves. New York: Springer-Verlag, 1992. (Undergrad. Texts in Math.).

[30] *Whiteman A. L.* Cyclotomy and Jacobsthal sums // Amer. J. Math. 1952. Vol. 74, № 1. P. 89–99.

[31] *Zagier D.* A one-sentence proof that every prime $p \equiv 1 \pmod 4$ is a sum of two squares // Amer. Math. Monthly. 1990. Vol. 97, № 2. P. 144.

[32] https://ru.wikipedia.org/wiki/Алгоритм_Чиполлы

[33] https://en.wikipedia.org/wiki/Counting_points_on_elliptic_curves

[34] https://ru.wikipedia.org/wiki/Алгоритмы_быстрого_возведения_в_степень

[35] https://ru.wikipedia.org/wiki/Алгоритм_Шуфа

[36] https://ru.wikipedia.org/wiki/Алгоритм_Тонелли-Шенкса

[37] https://pari.math.u-bordeaux.fr/dochtml/html/Elliptic_curves.html

Николай Николаевич Осипов, Сибирский федеральный университет, Красноярск

nnosipov@rambler.ru

Множественная сложность построения правильного многоугольника

Е. С. Коган

Эта работа иллюстрирует важный метод на примере решения алгоритмической задачи.

Будем рассматривать следующую операцию: к подмножеству $A \subset \mathbb{C}$, содержащему числа x, y, добавляется любое из чисел $x + y$, $x - y$, xy, или (если $y \neq 0$) x/y, или такое z, что $z^2 = x$.

Основная теорема. *Пусть p — простое число Ферма, т. е. простое число вида $2^m + 1$, где m — степень двойки с натуральным показателем, ε — первообразный корень степени p из единицы*:

$$\varepsilon := \cos \frac{2\pi}{p} + i \sin \frac{2\pi}{p}.$$

Тогда из $\{1\}$ можно получить некоторое множество, содержащее корни p-й степени из единицы: $1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{p-1}$, *за $12p^2$ добавлений, определённых выше*.

Назовём сложность, рассмотренную в основной теореме, *множественной сложностью*. Такое понятие сложности отличается от сложности как времени работы алгоритма, находящего корни степени p из 1. Однако последняя сложность также пропорциональна p^2. Об алгоритмах вычисления корней p-й степени из 1 см. [4], а также [5]. О строгих определениях различных понятий сложности см. [1]. Автор не исследовал соотношение введённого понятия множественной сложности с этими определениями. В любом случае основная теорема не претендует на новизну.

Замечание 1. Из основной теоремы можно вывести следующее утверждение.

Пусть p — простое число Ферма. Тогда существует такое действительное число C, не зависящее от p, что из единичного отрезка можно получить правильный p-угольник за $C \cdot p^2$ операций проведения окружности

с центром в одной точке и проходящей через другую и проведения прямой через две точки[1)].

Замечание 2. Из основной теоремы также можно вывести её вещественный аналог, который состоит в следующем.

Существует такое число C, что для любого простого числа Ферма p число $\cos(2\pi/p)$ можно получить из $\{1\}$ за $C \cdot p^2$ операций, аналогичных определённым выше, причём корни извлекаются только из положительных чисел.

Доказательство основной теоремы проводится аналогично [2, п. 5.3.4, с. 83–88]. Оценка же, получающаяся из доказательства построимости, приведённого в [2, конец п. 5.3.3, с. 83], пропорциональна p^3.

Идея доказательства основной теоремы для $p = 5$. Сначала выразим через радикалы некоторые многочлены от ε.

Заметим, что $(\varepsilon + \varepsilon^4)(\varepsilon^2 + \varepsilon^3) = \varepsilon + \varepsilon^2 + \varepsilon^3 + \varepsilon^4 = -1$. Обозначим $T_0 := \varepsilon + \varepsilon^4$ и $T_1 := \varepsilon^2 + \varepsilon^3$. Тогда по теореме Виета T_0 и T_1 являются корнями уравнения $t^2 + t - 1 = 0$. Отсюда можно выразить T_0 (и T_1). Поскольку $\varepsilon \cdot \varepsilon^4 = 1$, по теореме Виета числа ε и ε^4 являются корнями уравнения $t^2 - T_0 t + 1 = 0$. Отсюда можно выразить ε (и ε^4).

Идея доказательства основной теоремы в общем случае. Сначала хорошо бы разбить сумму $\varepsilon + \varepsilon^2 + \ldots + \varepsilon^{p-1} = -1$на два слагаемых T_0, T_1, которые можно получить добавлениями, описанными в начале статьи (иными словами, *сгруппировать* удачным образом корни уравнения $1 + x + x^2 + \ldots + x^{p-1} = 0$). Затем нужно разбить каждую сумму T_k на два слагаемых $T_{k,0}, T_{k,1}$, которые можно получить такими добавлениями. И так далее, пока не получим $T_{\underbrace{1,\ldots,1}_{s}} = \varepsilon$.

Теорема о первообразном корне, приведённая, например, в [2, п. 5.3.3, с. 82], позволяет закодировать ненулевые вычеты по модулю p вычетами по модулю $p-1$. А именно, выбрав первообразный корень g по модулю p, мы вычет k по модулю $p-1$ сопоставляем (ненулевому) остатку от деления g^k на p. Это кодирование использовано в группировках, построенных выше для $p = 5$.

Введём определения и обозначения, необходимые для доказательства.

Множество $A \subset \mathbb{C}$ *построимо за* n *операций из множества* $B \subset \mathbb{C}$, если какое-нибудь множество $A' \supset A$ можно получить из B за n добавлений, описанных в начале статьи. Используем следующие обозначения:

[1)] Известна история об аспиранте, который разработал построение правильного многоугольника с 65 537 сторонами за 20 лет (см. [3, с. 43]).

- $p = 2^m + 1$ — простое число Ферма;
- $q_k = 2^{k+1}$ $(k = 0, 1, \ldots, m)$;
- $s_k = 2^{m-k-1} - 1$ $(k = 0, 1, \ldots, m)$;
- ε — первообразный корень степени p из единицы;
- g — первообразный корень по модулю p;
- $T_{k,r} := \sum_{a=0}^{2^{m-k}-1} \varepsilon^{g^{2^k \cdot a + r}}$ для каждого $k \in \{0, 1, \ldots, m\}$, $r \in \mathbb{Z}_{2^m}$.
  В частности, $T_{0,0} = -1$, а $T_{m,r} = \varepsilon^{g^r}$. Также $T_{k,r_1} = T_{k,r_2}$ при $r_1 \equiv r_2 \pmod{2^k}$, поэтому это определение осмысленно и при $r \in \mathbb{Z}_{2^k}$;
- $N_{k,t}$ — число пар (c, d) вычетов по модулю 2^{m-k-1}, $k \in \{0, 1, \ldots, m-1\}$, удовлетворяющих сравнению

$$g^{q_k \cdot c + t} + g^{q_k \cdot d + 2^k + t} \equiv 1 \pmod{p}, \quad t \in \mathbb{Z}_{2^m}. \tag{1}$$

  Числа $T_{k,r}$ и $N_{k,t}$ зависят от m, но, поскольку m зафиксировано, оно не указывается.

Лемма 1. *Для любых вычетов $t_1, t_2 \in \mathbb{Z}_{2^m}$, сравнимых по модулю 2^k, $k \in \{0, 1, \ldots, m-1\}$, верно равенство $N_{k,t_1} = N_{k,t_2}$.*

Доказательство. Достаточно показать, что $N_{k,t} = N_{k,2^k+t}$. Для этого сопоставим каждому решению (c, d) сравнения (1) пару $(d, c-1)$. Эти пары дают все решения сравнения (1) при замене t на $2^k + t$, поскольку следующие сравнения равносильны:

$$g^{q_k \cdot c + t} + g^{q_k \cdot d + 2^k + t} \equiv 1 \pmod{p},$$
$$g^{q_k \cdot d + (2^k + t)} + g^{q_k \cdot (c-1) + 2^k + (2^k + t)} \equiv 1 \pmod{p}. \qquad \square$$

Лемма 2. *Для любых $k \in \{0, 1, \ldots, m-2\}$ и $r \in \mathbb{Z}_{2^m}$*

$$T_{k+1,r} T_{k+1,2^k+r} = \sum_{s=0}^{2^k-1} N_{k,r-s} T_{k,s}.$$

Доказательство. Имеем

$$T_{k+1,r} T_{k+1,2^k+r} = \left(\sum_{c=0}^{s_k} \varepsilon^{g^{q_k \cdot c + r}} \right) \cdot \left(\sum_{d=0}^{s_k} \varepsilon^{g^{q_k \cdot d + 2^k + r}} \right) =$$
$$= \sum_{c=0}^{s_k} \sum_{d=0}^{s_k} \varepsilon^{g^{q_k \cdot c + r} + g^{q_k \cdot d + 2^k + r}} \overset{(*)}{=} \sum_{s=0}^{2^m-1} N_{k,r-s} \varepsilon^{g^s} \overset{(**)}{=} \sum_{s=0}^{2^k-1} N_{k,r-s} T_{k,s}. \tag{2}$$

Равенство $(*)$ из (2) доказывается группировкой одинаковых слагаемых. Действительно, сравнение

$$g^{q_k \cdot c + r} + g^{q_k \cdot d + 2^k + r} \equiv g^s \pmod{p}$$

равносильно сравнению (1) для $t = r - s$. А сравнение

$$g^{q_k \cdot c + r} + g^{q_k \cdot d + 2^k + r} \equiv 0 \pmod{p}$$

не имеет решений, поскольку оно равносильно следующим:

$$\begin{aligned} g^{q_k \cdot c + r} &\equiv (-1) \cdot g^{q_k \cdot d + 2^k + r} \pmod{p}, \\ g^{q_k \cdot c + r} &\equiv g^{2^{m-1}} \cdot g^{q_k \cdot d + 2^k + r} \pmod{p}, \\ q_k \cdot c + r &\equiv 2^{m-1} + (q_k \cdot d + 2^k + r) \pmod{2^m}, \\ q_k \cdot (c - d) &\equiv 2^{m-1} + 2^k \pmod{2^m}, \\ 2(c - d) &\equiv 2^{m-k-1} + 1 \pmod{2^{m-k}}. \end{aligned}$$

Последнее сравнение не имеет решений, так как в левой части стоит чётное число, а в правой — нечётное (2^{m-k-1} чётно, так как $k \leqslant m-2$).

Равенство $(**)$ из (2) получается из леммы 1 группировкой одинаковых $N_{k,r-s}$. □

Замечание. При $k = m-1$ произведение $T_{k+1,r} T_{k+1,2^k+r}$ равно

$$T_{m,r} T_{m,2^{m-1}+r} = \varepsilon^{g^r} \cdot \varepsilon^{g^{2^{m-1}+r}} = \varepsilon^{g^r} \cdot \varepsilon^{-g^r} = 1.$$

Лемма 3. *Для любого целого числа k от 0 до $m-1$ множество*

$$A = \{0, 1, \ldots, p\} \cup \{T_{k+1,r} \mid r \in \mathbb{Z}_{2^{k+1}}\}$$

построимо за $11 \cdot 4^k$ операций из множества $B = \{0, 1, \ldots, p\} \cup \{T_{k,r} \mid r \in \mathbb{Z}_{2^k}\}$.

Доказательство. Во-первых, для любых $k \in \{0, 1, \ldots, m\}$ и $t \in \mathbb{Z}_{2^m}$ выполняется $N_{k,t} \leqslant p$, так как в сравнении (1) одному вычету c может соответствовать не больше одного вычета d. Следовательно, все $N_{k,t}$ содержатся в B.

Докажем, что множество $P := \{T_{k+1,r} T_{k+1,2^k+r} \mid r \in \mathbb{Z}_{2^k}\}$ построимо из B меньше чем за $2 \cdot 4^k$ операций; в определении P вычет r берётся по модулю 2^k, а не 2^{k+1}, так как $T_{k+1,r} T_{k+1,2^k+r} = T_{k+1,2^k+r} T_{k+1,2^k+(2^k+r)}$.

Из замечания к лемме 2 следует, что $P = \{1\} \subset B$ при $k = m-1$. Если же $k \leqslant m-2$, то множество $P' := \{N_{k,r-s} T_{k,s} \mid r, s \in \mathbb{Z}_{2^k}\}$ построимо из B за $2^k \cdot 2^k = 4^k$ операций умножения (можно для всех пар (r, s) добавить $N_{k,r-s} T_{k,s}$), а множество P по лемме 2 построимо из P' за $(2^k - 1) \cdot 2^k < 4^k$ операций умножения. Значит, множество P построимо из B меньше чем за $4^k + 4^k = 2 \cdot 4^k$ операций.

Далее, множество $P \cup B$ содержит

$$T_{k+1,r} + T_{k+1,2^k+r} = T_{k,r} \in B \quad \text{и} \quad T_{k+1,r} T_{k+1,2^k+r} \in P,$$

и для любых (комплексных) чисел x_1, x_2 множество $\{x_1, x_2\}$ построимо за 9 операций из множества $\{x_1 + x_2, x_1x_2\}$ (по формуле корней квадратного уравнения). Следовательно, $A' := \{T_{k+1,r} \mid r \in \mathbb{Z}_{2^{k+1}}\}$ построимо из $P \cup B$ за $9 \cdot 2^k$ операций, т. е. A' построимо из B за $2 \cdot 4^k + 9 \cdot 2^k \leqslant 11 \cdot 4^k$ операций. Кроме того, $\{0, 1, \ldots, p\} \subset B$, поэтому A также построимо из B за $11 \cdot 4^k$ операций. □

Доказательство основной теоремы. Из леммы 3 следует, что множество

$$\{T_{m,r} \mid r \in \{0, 1, \ldots, 2^m - 1\}\} = \{\varepsilon^r \mid r \in \{0, 1, \ldots, 2^m - 1\}\}$$

построимо из $\{1\}$ за

$$p + 1 + \sum_{k=0}^{m-1} 11 \cdot 4^k = p + 1 + 11 \cdot \frac{4^m - 1}{4 - 1} < p + 11 \cdot 4^m < 12p^2$$

операций. □

Благодарности

Благодарю Д. Мусатова, А. Савватеева и руководителя работы А. Скопенкова за ценные замечания и предложения при написании данной работы.

Список литературы

[1] *Абрамов С. А.* Лекции о сложности алгоритмов. М.: МЦНМО, 2012.

[2] *Заславский А. А., Скопенков А. Б., Скопенков М. Б.* Элементы математики в задачах. М.: МЦНМО, 2018. С. 82–90.

[3] *Литлвуд Дж.* Математическая смесь. М.: Физматлит, 1990.

[4] *Сафин А. Р.* Программа для построения правильных многоугольников циркулем и линейкой. https://www.mccme.ru/mmks/dec08/Safin.pdf

[5] *Berndt B., Evans R., Williams K.* Gauss and Jacobi sums. New York: John Wiley & Sons, Inc., 1998. (Canadian Mathematical Society Series of Monographs and Advanced Texts).

Евгений Сергеевич Коган, ученик 11 класса ГБОУ «Интеллектуал»
koganeser@gmail.com

Популяризация математики

О Санкт-Петербургской заочной олимпиаде по топологии

Н. С. Калинин

§ 1. Об организации олимпиады

Санкт-Петербургская заочная олимпиада по топологии проходила в 2017 и 2018 годах в октябре. Причина, по которой я стал её организовывать, проста: будучи студентом, я хотел поучаствовать в заочной олимпиаде по топологии, а её не было. Задачи олимпиады (приведённые ниже) я подбирал из эстетических и педагогических соображений: короткие формулировки, отсутствие формул, возможность «крутить» в голове картинки, необычный взгляд на стандартные объекты, возможность выучить что-то новое, — ведь олимпиада проходит целый месяц, можно пару книжек изучить.

Думаю, что задачи олимпиады были доставлены всем желающим. Мой пост «В контакте» в 2017 году посмотрели 13 000 человек, паблик «В контакте» sci-hub (Александра Элбакян любезно согласилась порекламировать) — 16 000; в фейсбуке, наверное, порядок величины такой же (по крайней мере пятьдесят человек поделились записью), 8 тыс. просмотров на quora получил пост незнакомого мне человека. Энтузиасты быстро перевели условия на португальский, английский, испанский, итальянский, французский, турецкий, потом и фарси добавился. Так что тысяч сто человек увидели условия, тысяча открыла файл, сто прочитали, сорок порешали, двадцать написали решения. Хотя и профессора решали задачи с большим удовольствием — говорили, что задачи красивые. В 2018 году просмотров в социальных сетях было больше, переводов меньше.

В 2017 году на русском я получил двенадцать работ, некоторые от команд из трёх человек, одну работу на испанском. В английской версии в правилах было написано, что работы будут проверены только при наличии свободного времени у организаторов. В 2018 году я получил решения от шести команд (все на русском).

Почти все участники — из Санкт-Петербурга, Москвы и Новосибирска. Разумно было предположить, что они узнали об олимпиаде от преподавателей (в 2017 я написал письма, кажется, во все университеты, где люди понимают русский, послал олимпиаду на пару сотен адресов, человек двадцать печатали и вешали олимпиаду на стенке). Но три четверти приславших решения узнали про олимпиаду из «В контакте» или фейсбука.

Так я себе представляю популяризацию науки, топологии в частности. Через вовлечение в решение задач. Будет замечательно, если такие олимпиады будут проведены по теории чисел, комбинаторике и так далее. Например, каждый год — такая заочная олимпиада по какой-то области математики, лучше в октябре, когда студенты уже вошли в ритм учебной жизни, но ещё ею не перегружены и до экзаменов-зачётов далеко. Возможные критерии трудности задач — одна-две задачи, которые может решить средний третьекурсник, и другие задачи, которые должны быть в первую очередь интересны и раскрывать темы, слабо отражённые в программе. Олимпиада не должна ничего проверять, должна стимулировать к изучению нового (как и все заочные олимпиады). Очень полезно давать на олимпиаду нерешённые задачи (и об этом предупреждать). Известные задачи тоже можно давать, потому что соревновательный элемент сведён практически к нулю, и победа в олимпиаде не должна давать никакого бюрократического или репутационного преимущества.

Организовывать олимпиаду мне интересно (не обязательно по топологии, можно по теории чисел или комплексному анализу), но основная проблема в задачах: неясно, где их брать.

§ 2. Правила

1. Олимпиада проходит с 1 октября по 31 октября включительно. Допускаются команды из 1–3 человек. Разрешается пользоваться любыми материалами. Запрещается просить помощи у кого-то, кроме участников своей команды. Призов не будет. Не у всех задач известны решения. Условия задач (актуальная версия, со всеми уточнениями по условиям, если они появятся) и полная версия правил были доступны по ссылке `http://mathcenter.spb.ru/nikaan/olympiad/`

2. Работы принимаются до 23:59 31 октября (по московскому времени).

3. Принимаются работы только в формате pdf: или набранные и скомпилированные в TEX, или написанные **понятным** почерком и отсканированные в **хорошем** качестве.

4. Участвовать в олимпиаде могут все. Отдельный зачёт проводится среди студентов 1–4 курсов. В работе указывайте, как вас зовут и где вы учитесь или работаете. Посылая работу, вы тем самым даёте согласие на обработку этих персональных данных (потом на сайте появится сводная таблица с результатами). Если это чем-то не устраивает, отдельно напишите — анонимные работы тоже допускаются, хотя и не приветствуются.

5. Если вы решили хотя бы одну задачу, заполните, пожалуйста, листок анонимного опроса
`https://docs.google.com/forms/d/1FUH6-JbcjSIIVMKMwxQzUpIW25pfNSPHEYjc2-WnIrE/`
Присылайте ваши любимые задачи по топологии для последующих олимпиад.

6. Результаты появятся на `http://mathcenter.spb.ru/nikaan/olympiad.html`

7. Приветствуется распространение задач олимпиады. Переводы задач на другие языки будут расположены по адресу
`http://mathcenter.spb.ru/nikaan/olympiade.html`

§ 3. Условия задач 2017 года

1. Найдите два таких негомеоморфных компактных подмножества X_1, X_2 плоскости, что $X_1 \times I$ гомеоморфно $X_2 \times I$, где $I = [0, 1]$ — замкнутый отрезок прямой.

2. Четырёхточечное множество наделили минимальной (по количеству открытых подмножеств) топологией, в которой две точки открыты (каждая из них является открытым множеством), а остальные две — замкнуты. Вычислите фундаментальную группу этого пространства и постройте его универсальное накрытие.

3. а) В пространстве всех треугольников на плоскости является ли деформационным ретрактом подпространство всех прямоугольных треугольников?

б) Построить деформационную ретракцию пространства всех треугольников плоскости на подпространство правильных треугольников.

4. Пусть X — связное многообразие и $f : X \to S^2$ — такое непрерывное отображение, что $f^{-1}(x)$ гомеоморфно S^1 для всех точек x сферы S^2. Чему может равняться $H_1(X, \mathbb{Z}), H_2(X, \mathbb{Z})$?

5. Рассмотрим $S^4 = \{x \in \mathbb{R}^5 \mid |x| = 1\}$. Можно ли в каждой точке $x \in S^4$ так выбрать двумерную аффинную плоскость P_x, касающуюся S^4 в x, что P_x непрерывно зависит от x?

6. Может ли хаусдорфово пространство со счётным количеством точек быть связным?

7. Существует ли сюръективное непрерывное отображение $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, не являющееся гомеоморфизмом, но такое, что у каждой точки $x \in \mathbb{R}^2$ существует такая окрестность U, что $f\colon U \to f(U)$ — гомеоморфизм?

8. а) Рассмотрим квадрат $[0, n]^2$ на плоскости, n — натуральное. Удалим из квадрата все точки, у которых обе координаты нецелые. Остался одномерный клеточный комплекс, назовём его X. Найдите максимальное такое $k = k(n)$, что для любого непрерывного отображения X в $\mathbb{R}^1$ найдётся точка с хотя бы k прообразами.

б) То же самое для отображений в $\mathbb{R}^2$ двумерного комплекса, полученного из $[0, n]^3 \subset \mathbb{R}^3$ выбрасыванием всех точек, у которых все координаты нецелые.

9. Существует ли такая иммерсия сферы $f\colon S^2 \to \mathbb{R}^3$, что не существует иммерсии $g\colon D^3 \to \mathbb{R}^3$ трёхмерного диска, для которой $g|_{\partial D^3} = f$?

§ 4. Условия задач 2018 года

1. а) Пусть $f\colon \mathbb{R} \to \mathbb{R}^2$ — инъективное непрерывное отображение, неограниченное на положительной и отрицательной полупрямых. Может ли его образ иметь связное дополнение?

б) Пусть $f\colon \mathbb{R} \to \mathbb{R}^3$ — инъективное непрерывное отображение, образ которого всюду плотен. Может ли дополнение его образа быть односвязным (иными словами, может ли быть так, что любое непрерывное отображение $S^1 \to \mathbb{R}^3 \setminus f(\mathbb{R})$ продолжается до непрерывного отображения из двумерного диска в $\mathbb{R}^3 \setminus f(\mathbb{R})$)?

2. а) Обозначим единичный открытый интервал $(0, 1) \subset \mathbb{R}$ через I. Пусть $f\colon I^2 \to \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ — диффеоморфизм. Докажите, что существует такая точка $x \in I$, что кривая $f(x \times I)$ имеет длину хотя бы 2. (Иными словами: диск гладко расслоён на кривые. Докажите, что какая-то из этих кривых не короче диаметра диска.)

б) Пусть $f\colon I^3 \to \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 < 1\}$ — диффеоморфизм. Получите как можно лучшую нижнюю оценку на супремум площадей множеств вида $f(x \times I^2)$, где $x \in I$.

3. Существует ли гладкое отображение $\mathbb{R}P^2 \to M$, имеющее регулярное значение с нечётным числом прообразов, если а) $M = S^2$; б) $M = S^1 \times S^1$?

в) Пусть M — связное замкнутое ориентированное 4-многообразие. Пусть $f\colon \mathbb{C}P^2 \to M$ — непрерывное отображение ненулевой степени. Найдите все возможные значения второго числа Бетти многообразия M. Напомним, что *второе число Бетти*, $b_2(M)$, равно наибольшему такому r,

что существует r замкнутых двумерных подмногообразий $X_1, \dots, X_r$ многообразия M и ранг матрицы алгебраических пересечений $X_i \cdot X_j$ равен r.

4. Один и тот же граф G нарисован двумя способами G_1, G_2 на единичной сфере $S^2 \subset \mathbb{R}^3$ так, что каждый из рисунков G_i центрально-симметричен относительно центра сферы и никакое ребро рисунка не пересекает другое ребро во внутренней точке и не проходит через вершины графа. Верно ли, что графы, полученные факторизацией по центральной симметрии S^2 из графов G_1, G_2, изоморфны?

(Пусть у двух графов, вложенных в $\mathbb{R}P^2$, прообразы при двулистном отображении $S^2 \to \mathbb{R}P^2$ изоморфны. Правда ли, что эти графы изоморфны?)

5. а) Докажите, что из любых 11 точек в трёхмерном пространстве можно выбрать три попарно непересекающиеся тройки так, чтобы три треугольника, образованные этими тройками, имели общую точку.

б) Верно ли аналогичное утверждение для 10 точек?

6. Пусть K — (двумерный) многоугольник на плоскости и a — вектор, причём образ $K + a$ многоугольника K при сдвиге на вектор a не пересекается с K, т. е. $K \cap (K + a) = \varnothing$. Докажите, что два воза, т. е. два круга диаметром $|a|$, не могут поменяться местами при непрерывном движении их центров по K, не столкнувшись.

7. Пусть $\gamma\colon [0, 1] \to \mathbb{R}^2$ — гладкая параметризованная кривая, у которой кривизна в точке $\gamma(t)$ монотонно убывает по t.

а) Покажите, что соприкасающиеся окружности в разных точках кривой не пересекаются.

б) Сейчас мы приведём цепочку рассуждений, приводящих к противоречию. Необходимо найти неверный шаг в этой цепочке.

> Заметим, что окрестность точки $\gamma(1/2)$ расслоена на соприкасающиеся окружности в точках кривой, близких к $\gamma(1/2)$. Тем самым кривая γ касается всех слоёв этого локального расслоения. Но это невозможно, так как локально это расслоение является произведением отрезков и если бы γ касалась всех слоёв, у неё бы везде была производная нуль, и значит, она бы локально совпадала с одним из слоёв, что неверно.

§ 5. Подсказки 2017 года

1. *Пример*: X_1 — кольцо ($\{z \in \mathbb{C} \mid 1 \leqslant |z| \leqslant 2\}$) с двумя волосками (отрезками) наружу с внешней окружности. Множество X_2 — такое же кольцо на плоскости с двумя волосками, расположенными по-другому: один волосок наружу с внешней окружности, второй вовнутрь с внутренней окружности.

2. Надо заметить, что это четырёхточечное множество по существу является окружностью (есть разбиение окружности на четыре клетки: две точки и два интервала). Фундаментальная группа, тем самым, есть $\mathbb{Z}$. Универсальным накрытием окружности является прямая, а у нас получится $\mathbb{Z}$ с топологией «цифровой прямой» (прямая Халимского).

3. а) *Ответ*: нет.

У прямоугольных треугольников есть выделенная вершина. Поэтому если есть такая ретракция, то у каждого треугольника можно каким-то образом однозначно выбрать вершину (которая пойдёт в прямой угол). Но рассмотрим правильный треугольник. В пространстве всех треугольников есть петля, которая представляет поворот на $2\pi/3$, и тогда правильный треугольник переходит в себя. Уже для этой петли нет ретракции.

б) Воспользуйтесь треугольником Наполеона.

4. Если f — расслоение, можно написать спектральную последовательность. Или можно считать, что S^2 склеено из двух дисков D_1, D_2 по их общей окружности $\partial D_1 = \partial D_2$. Значит, расслоения получаются как склейка $D_1 \times S^1$, $D_2 \times S^1$. С точностью до гомотопии важно, сколько раз «прокрутилась» вторая координата $x \times S^1$, пока точка x пробегала границу D_1. Получается отображение из S^1 в S^1, таких отображений с точностью до гомотопии $\pi_1(S^1)$, т. е $\mathbb{Z}$. Гомологии посчитать несложно при такой явной конструкции.

Но f необязательно расслоение! Бывают слоения Зейферта.

5. *Ответ*: нет.

Пусть такое подрасслоение есть. Тогда $H_2(S^4) = 0$ (класс Эйлера нулевой). Значит, и класс Эйлера суммы слоения и ему ортогонального равен нулю, а это не так: класс Эйлера касательного расслоения к S^4 равен 2.

6. *Ответ*: да, есть множество примеров. Ищите в интернете (на английском языке).

7. *Ответ*: да.

Плоскость $\mathbb{R}^2$ гомеоморфна длинному прямоугольнику без границы и кругу без границы (радиуса чуть больше, чем меньшая сторона прямоугольника). Уложим прямоугольник следующим образом: когда он первый раз проходит по кругу, он накрывает всё, кроме маленькой окрестности центра (похоже на отображение $z \to e^z$ на комплексной плоскости), а при проходе второй раз мы уложим конец прямоугольника на центр круга.

8. а) Можно посмотреть на образы вершин. Если они все различны, то рассмотрим прообраз точки, у которой примерно поровну образов вершин справа и слева.

б) Я умею получать неоптимальную оценку (типа $n/10$).

9. *Ответ*: конечно.

Давайте сначала посмотрим на плоскость — вложим окружность восьмёркой. Понятно, что это невозможно продолжить иммерсией диска. Такую же «восьмёрку» сделаем в пространстве: рассмотрим стандартную сферу, вдавим верхнюю её часть, чтобы она оказалась ниже нижней. Нам надо всего лишь, чтобы образ вложения сферы делил пространство на части. Теперь посмотрим на возможную иммерсию D^3 без границы. Заметим, что это открытое связное множество. Значит, его граница не может быть трёхмерной «восьмёркой».

§ 6. Подсказки 2018 года

1. а) *Ответ*: можно.

Заметьте, что нет противоречия с леммой Жордана — нельзя по непрерывности добавить к нашей кривой точку «на бесконечности», замкнув кривую в окружность. В самом деле, параметризуем кривую временем t. Кривая может уходить далеко от начала координат, потом возвращаться всё ближе к началу координат (не заходя в него) и делать так бесконечное число раз при $t \gg 0$ и $t \ll 0$. Пусть можно разбить дополнение к кривой на два открытых множества. Посмотрим, в какую из частей попало начало координат, и придём к противоречию.

Линейной связности добиться невозможно (как заметили внимательные читатели), см. доказательство:
`http://mathcenter.spb.ru/nikaan/olympiad/topology2018error.pdf`

2. а) Посмотрите на кривую, проходящую через центр круга.

б) Ответ неизвестен (и участники олимпиады не предоставили никаких оценок).

3. а) *Ответ*: да.

б) *Ответ*: нет.

Опишем геометрическую идею. Возьмём меридиан m и параллель l в торе $S^1 \times S^1$. Заметим, что m и l пересекаются в одной точке, значит, их прообразы $m', l' \subset \mathbb{R}P^2$ тоже должны «пересекаться» в одной точке. А небольшой сдвиг m в $S^1 \times S^1$ не пересекается с m, то же и для l, значит, и их прообразы должны обладать таким же свойством. Легко видеть, что на $\mathbb{R}P^2$ нет кривых m', l' с заданными свойствами: если представить $\mathbb{R}P^2$ как ленту Мёбиуса, заклеенную диском, то каждая кривая в $\mathbb{R}P^2$ либо затягивается диском (и тогда может не пересекаться с некоторым своим сдвигом), либо не затягивается диском и тогда не отличается от центральной линии ленты Мёбиуса — но тогда любой её сдвиг пересекается с ней. Получается, что если l', m' пересекаются, то они должны быть по существу центральными линиями ленты Мёбиуса, но тогда они будут пересекаться

и с любыми своими шевелениями. Эту геометрическую идею можно формализовать, но проще выучить, что такое гомологии и когомологии (грубо говоря, это некоторые группы, которые кодируют пересечения подмногообразий нашего многообразия, в нашем случае пересечения кривых в поверхностях).

в) На когомологическом языке наличие отображения ненулевой степени гарантирует наличие отображения $f^*\colon H^*(M,\mathbb{Z}) \to H^*(\mathbb{C}P^2,\mathbb{Z})$. При этом $f^*(a\cdot b) = f^*(a)\cdot f^*(b)$. Отсюда несложно вывести, что $b_2(M) \leqslant \leqslant b_2(\mathbb{C}P^2) = 1$, а потом привести примеры.

4. Обозначим центральные симметрии сфер через σ_1, σ_2, они действуют на вершинах и рёбрах графа. Найдём такие две точки v, w, которые переходят одна в другую при обеих центральных симметриях, $\sigma_1(v) = = \sigma_2(v) = w$. Сделать это можно так: соединим вершину v графа кратчайшим путём γ с $\sigma_1(v)$. Рассмотрим, как действует σ_2 на $\gamma \cup \sigma_1(\gamma)$. Посчитайте количество вершин графа внутри петли $\gamma \cap \sigma_1(\gamma)$, их должно быть столько же, сколько и снаружи. Из этого следует, что $\gamma \cap \sigma_1(\gamma)$ пересекается с $\sigma_2(\gamma \cap \sigma_1(\gamma))$, отсюда следует существование искомых v, w, отсюда вытекает решение задачи. Некоторые хлопоты доставляет следующий факт: граф без v и $\sigma_1(v)$ может распасться на несколько компонент связности, и с этим нужно разбираться отдельно.

5. См. статьи: K. S. Sarkaria, «A generalized van Kampen — Flores theorem»[1] и А. Ю. Воловиков, «К теореме Ван-Кампена — Флореса»[2]. Интересно было бы узнать элементарное доказательство.

6. Если два воза могут поменяться местами, то в какой-то момент вектор, соединяющий их центры, параллелен a и длиннее $|a|$, причём целиком лежит в K, а значит, перенос K на a пересекается с K.

7. а) Прямые вычисления.

б) Получаемое «расслоение» не является гладким, поэтому понятия производной и, соответственно, касания теряют смысл.

[1] PAMS, 1991, vol. 111, № 2, p. 559–565.

[2] «Математические заметки», 1996, том 59, выпуск 5, с. 663–670.

Никита Сергеевич Калинин, НИУ ВШЭ (СПб)
nikaanspb@gmail.com

По мотивам задачника

Протыкание семейства транслятов двумерного выпуклого тела

Р. Н. Карасёв

§ 1. Формулировка результата

Бранко Грюнбаум в 1960-х годах поставил такую задачу: доказать, что для семейства транслятов (результатов параллельного сдвига) выпуклого тела на плоскости, в котором любые два множества пересекаются, существует 3-*трансверсаль*, т. е. такие три точки, что каждое множество семейства содержит хотя бы одну из них. Напомним, что *выпуклое тело* — это выпуклый компакт с непустой внутренностью.

В «Математическом просвещении» (сер. 3, вып. 19, с. 258) опубликована следующая достаточно наглядная задача 10, включающая задачу Грюнбаума. *На столе лежат круглые салфетки, возможно, разного размера. Любые две пересекаются. Докажите, что их можно прибить* 100 *гвоздями. Можно ли уменьшить число* 100? *А если эти салфетки суть единичные круги*? *А если это выпуклые фигуры, отличающиеся параллельным переносом*?

Следующие теоремы отвечают на эти вопросы.

Теорема 1.1. *Для семейства* $\mathscr{F} = \{K + x : x \in X\}$ *транслятов выпуклого тела* K *на плоскости* $\mathbb{R}^2$, *в котором любые два транслята пересекаются, найдётся множество* P *из не более чем трёх точек, пересекающее любой транслят.*

Теорема 1.2. *Для семейства* $\mathscr{F}$ *кругов в* $\mathbb{R}^2$, *в котором любые два круга пересекаются, найдётся множество* P *из не более чем* 12 *точек, пересекающее любой круг семейства.*

Константу 12 в последнем утверждении можно уменьшить до 4 (результат Данцера [1]), но доказательство такого результата непростое.

Дальнейший текст является слегка исправленной версией статьи [2]. Аналогичные этой более общие задачи рассматривались в [3].

§ 2. Вспомогательные факты

Далее запись $A-B$, где $A, B \subseteq \mathbb{R}^2$, обозначает множество всех векторов вида $a-b,\ a \in A,\ b \in B$. Переформулируем утверждение теоремы 1.1:

Теорема 2.1. *Если $X-X \subseteq K-K$ для множества $X \in \mathbb{R}^2$ и выпуклого тела K, то X можно покрыть тремя транслятами тела K.*

Лемма 2.2. *Теоремы* 1.1 *и* 2.1 *эквивалентны.*

Доказательство. Покажем, что теорема 1.1 следует из теоремы 2.1. Так как для любых точек $x_1, x_2 \in X$ оказывается $(K+x_1) \cap (K+x_2) \neq \varnothing$, то для любых $x_1, x_2 \in X$ найдётся такая точка p, что $p = x_1 + y_1$ и $p = x_2 + y_2$, где $y_1, y_2 \in K$, откуда $x_1 - x_2 = y_2 - y_1$, что и означает $X - X \subseteq K - K$.

По теореме 2.1 существуют такие точки x_1, x_2, x_3, что для любой точки $x \in X$ найдётся такое $y \in K$, что $x = x_i - y$ при некотором i. Это и требуется в теореме 1.1.

В обратную сторону утверждение доказывается теми же рассуждениями в обратном порядке. □

Для ограниченных множеств $X, K \subset \mathbb{R}^2$ отношение длин проекций X и K на некоторое направление a в плоскости будем называть *шириной X в данном направлении относительно K* и обозначать $w(X, K, a)$.

Выпуклую оболочку произвольного множества A будем обозначать $\operatorname{conv}(A)$. Очевидно, что

$$w(X, K, a) = 2w(X, K-K, a),$$
$$w(X-X, K, a) = 2w(X, K, a),$$
$$w(X, K, a) = w(X-X, K-K, a),$$
$$w(X, K, a) = w(\operatorname{conv}(X), K, a).$$

Следующая простая лемма проясняет геометрический смысл включения $X - X \subseteq K - K$.

Лемма 2.3. *Следующие утверждения равносильны*:
i) $X - X \subseteq K - K$; ii) $w(X, K, a) \leqslant 1$ *для любого* a.

Доказательство. (i) $\Rightarrow$ (ii) Следует из равенства

$$w(X, K, a) = w(X-X, K-K, a).$$

(ii) $\Rightarrow$ (i) Так как множества $X-X$ и $K-K$ центрально-симметричны с центром в 0 и $w(X-X, K-K, a) \leqslant 1$, то проекция $X-X$ на любое направление содержится в проекции выпуклого тела $K-K$. По «теореме Хана — Банаха на плоскости» (не лежащая в выпуклом теле точка отделяется от него прямой) отсюда следует, что $X-X \subseteq K-K$. □

Из леммы 2.3 следует, что множество X в теореме 2.1 можно считать выпуклым, так как условие (ii) леммы не меняется при замене X на $\operatorname{conv} X$.

Лемма 2.4. *Если K — выпуклое тело в $\mathbb{R}^2$, а $X \subset \mathbb{R}^2$ и $w(X, K, a) \leqslant 1/2$ для любого направления a, то существует такое выпуклое тело F, что $X \subseteq F$ и $w(F, K, a) = 1/2$ для любого направления a.*

Набросок доказательства. Рассмотрим множество всех таких выпуклых тел F, что $F \supseteq X$ и $w(X, K, a) \leqslant 1$ для любого a. Оно непусто, так как содержит $\operatorname{conv}(X)$. Из леммы Цорна следует (подробности её применения оставляем читателю), что среди таких F существует максимальное по включению.

Далее нужно доказать, что если в каком-то направлении ширина F относительно K осталась меньше единицы, то F не было максимальным. Мы оставляем последнее утверждение читателю, *предупреждая*, что оно не очевидно и начиная с размерности 3 просто неверно. □

Таким образом, теперь теорема 2.1 с помощью леммы 2.4 может быть выведена из своего частного случая:

Теорема 2.5. *Если $X,\ K \subset \mathbb{R}^2$ — выпуклые тела и $X-X=K-K$, то X можно покрыть тремя транслятами K.*

Условие $X-X=K-K$ равносильно тому, что X и K имеют одинаковую ширину в любом направлении (см. лемму 2.3). Теперь сформулируем лемму, которая будет играть основную роль в доказательстве теоремы 2.5:

Лемма 2.6. *Пусть $\triangle A_1B_1C_1$ образован серединами сторон $\triangle ABC$. Если прямая ℓ не пересекает $\triangle A_1B_1C_1$ и не параллельна ни одной из его сторон, то ℓ образует с некоторыми двумя сторонами $\triangle ABC$ треугольник площади, большей чем $S_{\triangle ABC}$.*

Доказательство. Достаточно рассмотреть два существенно различных случая:

1) Прямая ℓ не пересекает $\triangle ABC$ и точка A наиболее удалена от ℓ. Тогда, очевидно, ℓ образует с прямыми AB и AC треугольник большей площади, чем $S_{\triangle ABC}$.

2) Прямая ℓ пересекает стороны AB и AC и луч BC в точках F, E и D соответственно. Тогда из условия леммы следует, что $AE < EC$. Легко видеть,

что треугольник, симметричный треугольнику DEC относительно точки E, содержит $\triangle AFE$, а следовательно, $S_{\triangle AFE} < S_{\triangle DEC}$. Значит, $S_{\triangle BFD} > S_{\triangle ABC}$.

Остальные случаи сводятся к рассмотренным после переобозначения сторон треугольника. □

§ 3. Доказательство теоремы 1.1

Введём обозначения. Пусть a — вектор единичной длины, X — выпуклое тело. Обозначим через $\ell_+(a, X)$ и $\ell_-(a, X)$ опорные прямые к X, перпендикулярные a и такие, что если $\lambda_1 a \in \ell_-(a, X)$, $\lambda_2 a \in \ell_+(a, X)$, то $\lambda_2 - \lambda_1 > 0$, иначе говоря, $(a, x) > 0$ для всех $x \in \ell_+(a, X) - \ell_-(a, X)$.

Можно заметить, что $\ell_+(a, X) = \ell_-(-a, X)$. Заметим также, что сумма или разность двух параллельных прямых — прямая, параллельная им обеим. Тогда можно обозначить

$$m(a, X) = \frac{1}{2}(\ell_+(a, X) + \ell_-(a, X)),$$

т. е. прямая $m(a, X)$ равноудалена от $\ell_+(a, X)$ и $\ell_-(a, X)$ и параллельна им обеим, делит ширину X в заданном направлении пополам.

Доказательство теоремы 2.5. Сначала явно построим три транслята фигуры K, а потом докажем, что они покрывают X.

Пусть a — некоторый вектор длины 1. Рассмотрим прямую

$$\ell(a) = \ell_+(a, X) - \ell_+(a, K).$$

Так как $w(X, K, a) = 1$, имеем $\ell(a) = m(a, X - K)$. Ясно, что $\ell(a)$ непрерывно зависит от a. Для попарно неколлинеарных векторов a_1, a_2, a_3 обозначим через $S(a_1, a_2, a_3)$ площадь треугольника, образованного прямыми $\ell(a_1)$, $\ell(a_2)$, $\ell(a_3)$. Так как эти прямые пересекаются внутри $X - K$, то

$$S(a_1, a_2, a_3) \leqslant \frac{1}{2}(\operatorname{diam}(X - K))^2 \sin\varphi,$$

где φ — угол между прямыми $\ell(a_1)$ и $\ell(a_2)$. Следовательно, величина $S(a_1, a_2, a_3)$ стремится к нулю, когда какие-то из направлений $\ell(a_i)$ стремятся друг к другу, поэтому её можно считать непрерывной функцией набора a_1, a_2, a_3, в котором могут быть и равные векторы. По соображениям компактности S при некоторых a_1, a_2, a_3 достигает максимума.

Если этот максимум равен нулю, то любые три, а значит, и все $\ell(a)$ пересекаются в одной точке t. Тогда $X = K + t$, так как опорные прямые к ним в любом направлении совпадают. В самом деле,

$$t \in \ell_+(a, X) - \ell_+(a, K),$$

а следовательно,

$$0 \in \ell_+(a, X) - \ell_+(a, K) - t,$$

т. е.

$$0 \in \ell_+(a, X) - \ell_+(a, K + t) \quad \text{и} \quad \ell_+(a, K + t) = \ell_+(a, X).$$

В противном случае пусть t_1, t_2 и t_3 — середины соответствующих сторон треугольника, образованного прямыми $\ell(a_1)$, $\ell(a_2)$ и $\ell(a_3)$. Покажем, что трансляты $K_i = K + t_i$ покрывают X. Будем считать, что $(a_i, t_i - t_j) > 0$ $(i \neq j)$, меняя в случае необходимости знаки у a_i (см. рис. 1).

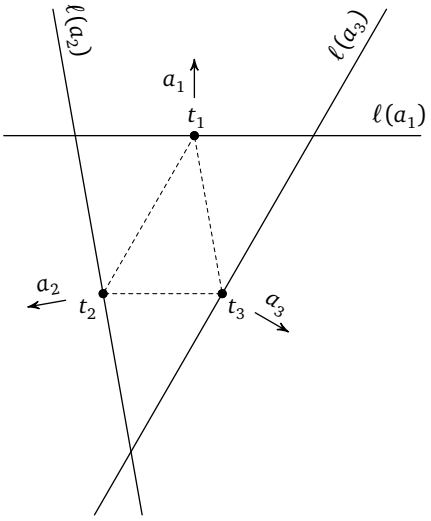


Рис. 1

Возьмём на границе K точки y_i, принадлежащие $\ell_-(a_i, K)$. Докажем, что $y_i + t_i - t_j \in K$, т. е. $y_i + t_i \in K + t_j$ $(i, j = 1, 2, 3)$. Достаточно доказать это, например, для $y_1 + t_1 - t_2$ и $y_1 + t_1 - t_3$, так как для других i доказательство аналогично.

На самом деле достаточно доказать, что выпуклая оболочка точки y_1 и точек касания $\ell_+(a_2, K)$ и $\ell_+(a_3, K)$ с K содержит y_1, $y_1 + t_1 - t_2$ и $y_1 + t_1 - t_3$ и содержится в K (это схематически изображено на рис. 2). Для этого достаточно доказать, что $\ell_+(a_2, K)$ и $\ell_+(a_3, K)$ не касаются K внутри полосы, образованной прямыми $\ell_-(a_1, K)$ и $\ell_-(a_1, K) + t_1 - t_2 = \ell_-(a_1, K) + t_1 - t_3$ (последняя прямая содержит точки $y_1 + t_1 - t_2$ и $y_1 + t_1 - t_3$, см. рис. 2; здесь используется тот факт, что прямая $t_2 t_3$ параллельна $\ell(a_1)$, а потому и $\ell_-(a_1, K)$).

Докажем, что точка касания $\ell_+(a_3, K)$ и K не лежит внутри полосы, образованной прямыми $\ell_-(a_1, K)$ и $\ell_-(a_1, K) + t_1 - t_2$. Для второй прямой

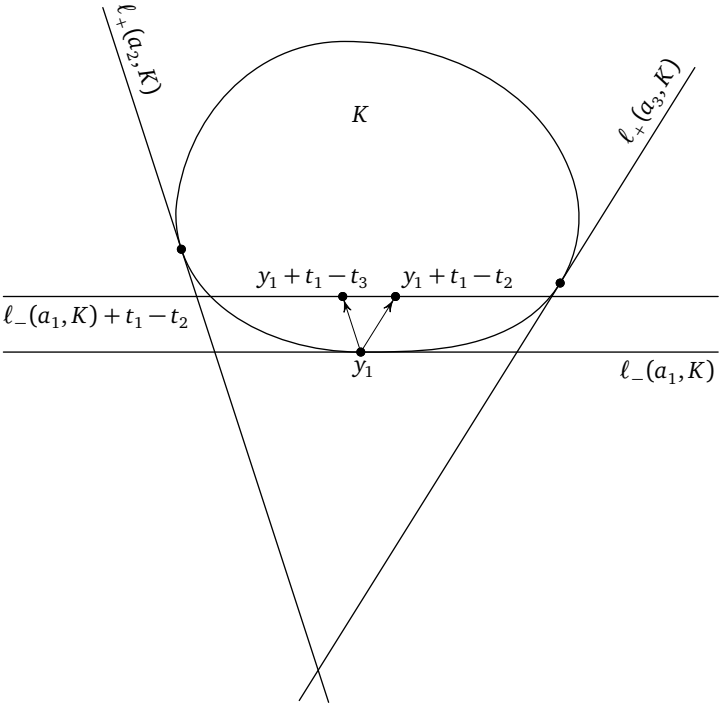


Рис. 2

доказательство аналогично. Сдвинем все рассматриваемые прямые и точки на вектор t_2. Тогда K перейдёт в K_2.

Теперь нам нужно доказать, что точка касания $\ell_+(a_3, K_2)$ и K_2 не лежит внутри полосы, образованной прямыми $\ell_-(a_1, K_2)$ и $\ell_-(a_1, K_2) + t_1 - t_2 = \ell_-(a_1, K) + t_1 = \ell_-(a_1, X)$ (последнее равенство следует из того, что $t_1 \in \ell(a_1) = \ell_-(a_1, X) - \ell_-(a_1, K)$).

Предположим противное. Пусть вектор a получается из a_3 малым поворотом в сторону a_1. Тогда при достаточно малом отличии a от a_3 прямые $\ell_+(a_3, K_2)$ и $\ell_+(a, K_2)$ будут пересекаться в полосе между $\ell_-(a_1, K_2)$ и $\ell_-(a_1, X)$ (это схематически изображено на рис. 3).

Рассмотрим теперь прямые $\ell_+(a_3, X)$ и $\ell_+(a, X)$. Прямая $\ell_+(a_3, X)$ получается из $\ell_+(a_3, K_2)$ сдвигом на вектор $t_3 - t_2$, так как $t_3 \in \ell_+(a_3, X) - \ell_+(a_3, K)$, т. е. $t_3 - t_2 \in \ell_+(a_3, X) - \ell_+(a_3, K_2)$.

Векторы a_1, a_2, a_3 выбраны так, чтобы $\ell(a_1)$, $\ell(a_2)$, $\ell(a_3)$ образовали треугольник максимальной площади и прямая $\ell(a)$ была не параллельна его сторонам. По лемме 2.6 она пересекает $\triangle t_1 t_2 t_3$ и в данном случае обязана пересекать стороны $t_2 t_3$ и $t_1 t_3$ (см. рис. 1). Так как

$$\ell(a) = \ell_+(a, X) - \ell_+(a, K) = \ell_+(a, X) - \ell_+(a, K_2) + t_2,$$

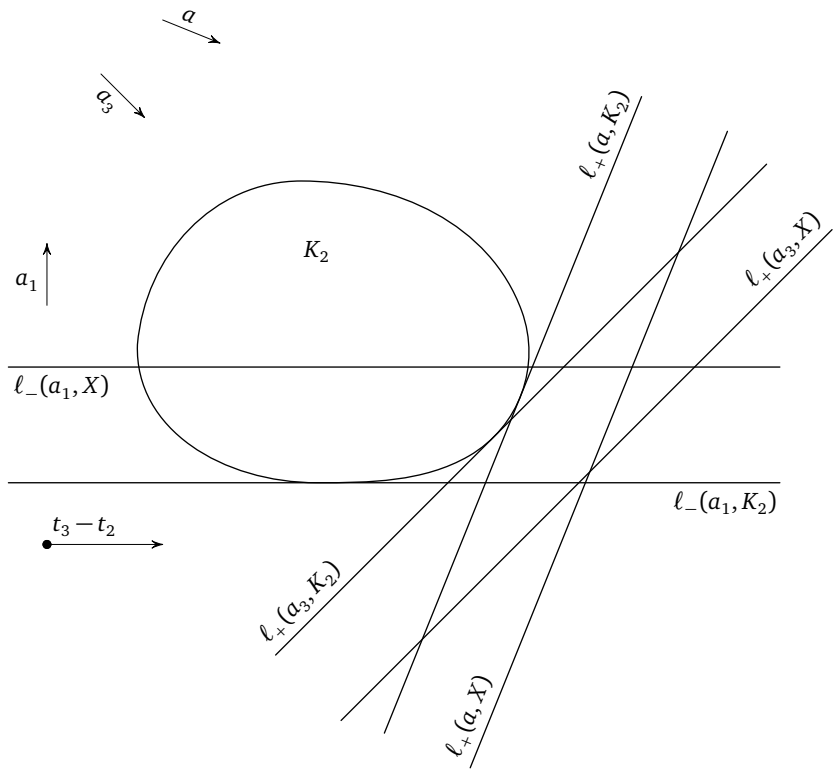


Рис. 3

прямая $\ell_+(a, X)$ получается из прямой $\ell_+(a, K_2)$ сдвигом на вектор, сонаправленный с $t_3 - t_2$ и меньшей длины.

Это означает, что точка пересечения $\ell_+(a_3, X)$ и $\ell_+(a, X)$ лежит по другую сторону от прямой $\ell_-(a_1, X)$ по сравнению с X, как и точка пересечения прямых $\ell_+(a_3, K_2)$ и $\ell_+(a, K_2)$ (см. рис. 3, эти сдвиги прямых $\ell_+(a_3, K_2)$ и $\ell_+(a, K_2)$ только удаляют точку их пересечения от $\ell_-(a_1, X)$). На рис. 3 также видно, что при таком расположении все три прямые $\ell_+(a, X)$, $\ell_+(a_3, X)$ и $\ell_-(a_1, X)$ не могут одновременно быть опорными для X. Получили противоречие.

Итак, по доказанному $y_i + t_i \in \bigcap_j K_j$ при $i = 1, 2, 3$. Пусть $C = \operatorname{conv} \bigcup_i K_i$, докажем, что $X \subseteq C$. Предположим противное и возьмём прямую, отделяющую какие-то точки X от C. Пусть a — её направляющий вектор, тогда прямые $\ell_+(a, X) - \ell_+(a, K_i)$ лежат по одну сторону от начала координат, т. е. все три точки t_i лежат по одну сторону от $\ell(a)$ — противоречие с леммой 2.6.

Множество C отличается от $\bigcup_i K_i$ на объединение множеств, лежащих внутри треугольников T_i, с вершинами $y_i + t_j$ ($j = 1, 2, 3$, см. рис. 4).

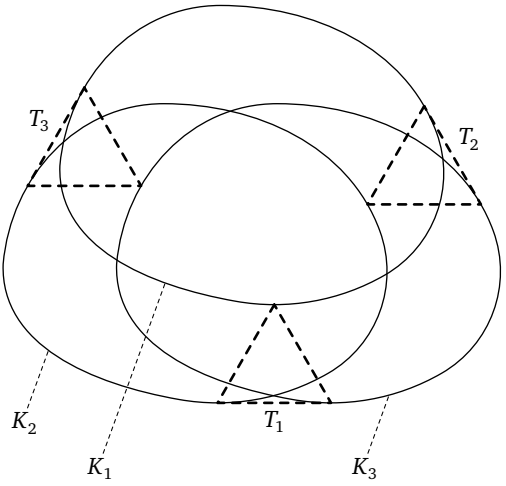


Рис. 4

Но $\triangle T_i$ лежит с X по разные стороны от $\ell\_(a_i, X)$, и поэтому множество $\triangle T_i \cap (C \setminus \bigcup_i K_i)$ не пересекает X. Значит, $C \setminus \bigcup_i K_i$ тоже не пересекает X. Следовательно, $X \subseteq \bigcup_i K_i$.

Теорема 2.5 доказана. □

§ 4. Набросок доказательства теоремы 1.2

Рассмотрим данную нам систему кругов.

1) Если мы рассмотрим минимальный круг в системе с радиусом r и нарисуем три равных круга не из системы радиусом

$$R = \frac{r}{\frac{2}{\sqrt{3}} - 1},$$

касающиеся внешним образом друг друга попарно и касающиеся маленького круга, то три точки их касания друг с другом протыкают все круги системы с радиусами, большими R. Для доказательства требуется заметить, что все круги системы пересекают маленький круг, и рассмотреть криволинейный треугольник, образованный частями трёх вспомогательных кругов радиуса R, и его пересечение с выбранным кругом системы.

2) Если мы рассмотрим максимальный по площади треугольник, который касается тремя своими сторонами снаружи некоторых трёх кругов из нашей системы, то точки его касания с кругами (по совместительству — его середины сторон) проткнут все круги системы, если отношение радиусов любых двух кругов в системе не больше 2. Это делается

с помощью леммы 2.6, которая показывает, что любой круг системы пересекает выбранный срединный треугольник, и некоторой «медитации» над картинкой. Если рассматриваемых в этих пунктах треугольников не существует, то любые три крута системы пересекаются и, значит, все они протыкаются общей точкой по теореме Хелли.

3) Из пункта (1) следует, что, проткнув слишком большие круги тремя точками, мы можем разбить оставшиеся круги на три системы с отношениями радиусов не более 2, так как

$$\frac{1}{\frac{2}{\sqrt{3}}-1} < 8,$$

и применить к этим системам приём (2). В итоге мы затратим не более 12 точек на протыкание всех кругов.

Список литературы

[1] *Danzer L.* Zur Lösung des Gallaischen Problems über Kreisscheiben in der Euklidischen Ebene // Studia Sci. Math. Hungar. 1986. Vol. 21, № 1–2. P. 111–134.

[2] *Карасёв Р. Н.* О трансверсалях семейств транслятов двумерного выпуклого компакта // Зап. научн. сем. ПОМИ. 1998. Т. 252. С. 67–77.

[3] *Karasev R. N.* Piercing families of convex sets with the d-intersection property in $\mathbb{R}^d$ // Discrete Comput. Geom. 2008. Vol. 39, № 4. P. 766–777.

Роман Николаевич Карасёв, МФТИ, ИППИ РАН

r_n_karasev@mail.ru

Задача о лягушке

А. Ю. Эвнин

Рассматривается олимпиадная задача, связанная с распределением двумерной случайной величины и допускающая разнообразные подходы к своему решению.

Начиная с 2009 г. в Южно-Уральском государственном университете регулярно проводятся математические конкурсы [1–5]. В числе участников этих конкурсов студенты и аспиранты ЮУрГУ, других вузов, а также любители математики из разных городов России и стран ближнего зарубежья.

В заметке разбирается задача из 55-го конкурса.

Задача. *Лягушка совершает прыжки, каждый — на метр. Направление каждого прыжка выбирается случайно* (*считаем, что случайная величина, равная углу поворота, распределена равномерно на отрезке* $[-\pi;\pi]$). *С какой вероятностью после трёх прыжков лягушка окажется на расстоянии не больше* 1 *м от начальной точки*?

§ 1. Неверное решение

Эта задача взята из материалов олимпиады AMC (American Mathematics Competitions) 2010 г. Авторы предложили такое решение.

Пусть первый прыжок лягушки был из точки A в точку O (рис. 1). Тогда после ещё двух прыжков лягушка может находиться в любой точке круга радиуса 2 с центром в точке O (будем называть этот круг *большим кругом*). Благоприятное событие состоит в попадании лягушки в *маленький круг* — круг единичного радиуса с центром в точке A. Отношение

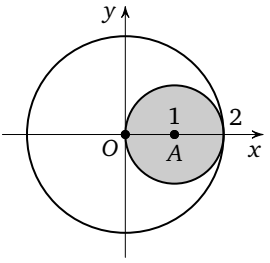


Рис. 1

Первоначальная версия опубликована в журнале «Математика в высшем образовании», 2018, № 16, с. 35–42.

площади маленького круга к площади большого, которое нас интересует, равно 1/4. Здесь получен верный ответ, но решение неверное!

Действительно, в нём неявно предполагается, что распределение положения лягушки после двух прыжков (если она начинает прыгать из точки O) является равномерным в большом круге. Основанием к такому выводу может служить следующее наблюдение: из точки O в любую точку внутри большого круга, отличную от его центра, можно за два прыжка попасть ровно двумя способами.

Однако указанное распределение не является равномерным. О том, каково оно, мы поговорим в конце статьи. А сначала решим задачу, не находя самого распределения.

§ 2. Три верных решения

Способ 1. Пусть лягушка прыгает по комплексной плоскости, стартуя из начала координат. Направим действительную ось в направлении её первого прыжка. Пусть направления второго и третьего прыжка составляют с действительной осью соответственно углы α и β. Можно считать, что α и β — случайные величины, равномерно распределённые на отрезке $[-\pi;\pi]$. Нас интересует вероятность выполнения неравенства

$$|1+e^{i\alpha}+e^{i\beta}|\leqslant 1.$$

Имеем:

$$(1+\cos\alpha+\cos\beta)^2+(\sin\alpha+\sin\beta)^2\leqslant 1;$$
$$3+2(\cos\alpha+\cos\beta+\cos(\alpha-\beta))\leqslant 1;$$
$$\cos\alpha+\cos\beta+\cos(\alpha-\beta)\leqslant -1;$$
$$2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}+2\cos^2\frac{\alpha-\beta}{2}-1\leqslant -1;$$
$$2\cos\frac{\alpha-\beta}{2}\left(\cos\frac{\alpha+\beta}{2}+\cos\frac{\alpha-\beta}{2}\right)\leqslant 0;$$
$$4\cos\frac{\alpha-\beta}{2}\cos\frac{\alpha}{2}\cos\frac{\beta}{2}\leqslant 0;$$

с вероятностью единица

$$\cos\frac{\alpha-\beta}{2}\leqslant 0.$$

Заметим, что, в наших предположениях, $-\pi\leqslant(\alpha-\beta)/2\leqslant\pi$. Поэтому окончательно имеем $-\pi\leqslant(\alpha-\beta)/2\leqslant-\pi/2$ или $\pi/2\leqslant(\alpha-\beta)/2\leqslant\pi$. Двумерная случайная величина $(\alpha/2;\beta/2)$ равномерно распределена на квадрате

$$D=\left[-\frac{\pi}{2};\frac{\pi}{2}\right]\times\left[-\frac{\pi}{2};\frac{\pi}{2}\right].$$

Найденным неравенствам между α и β (описывающим интересующее нас событие) соответствуют закрашенные треугольники на рис. 2.

Искомая вероятность равна отношению площади закрашенной области к площади квадрата D. Как видно, это $1/4$.

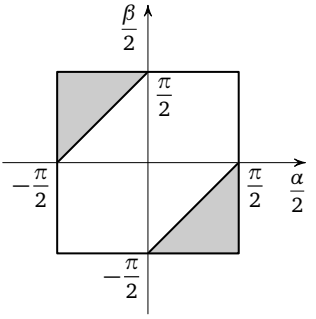


Рис. 2

Способ 2. Пусть лягушка прыгает по маршруту $A \to B \to C \to D$. Обозначим $\alpha = \angle ABC$, $\beta = \angle BCD$ (рис. 3). Удобно считать, что α и β — случайные величины, равномерно распределённые соответственно на отрезках $[0;\pi]$ и $[0;2\pi]$.

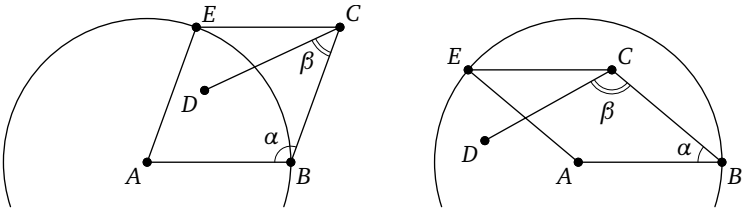


Рис. 3

Достроим треугольник ABC до ромба $ABCE$. Для того чтобы точка D оказалась в единичном круге с центром в точке A, необходимо и достаточно выполнение неравенства

$$0 \leqslant \beta \leqslant \pi - \alpha. \qquad (*)$$

Двумерная случайная величина $(\alpha;\beta)$ равномерно распределена на прямоугольнике $G = [0;\pi] \times [0;2\pi]$. А соотношению $(*)$ (т. е. интересующему нас событию) соответствует закрашенный треугольник на рис. 4.

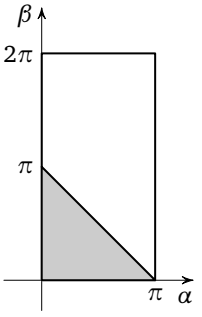


Рис. 4

Искомая вероятность равна отношению площади треугольника к площади прямоугольника. Как видно, это $1/4$.

Способ 3 (М. Д. Бронштейн, Э. Ю. Лернер).

Лемма[1]. *Пусть* $\boldsymbol{a}$, $\boldsymbol{b}$ *и* $\boldsymbol{c}$ *— компланарные, но попарно не коллинеарные единичные векторы. Тогда из векторов* $\boldsymbol{a}+\boldsymbol{b}+\boldsymbol{c}$, $\boldsymbol{a}-\boldsymbol{b}+\boldsymbol{c}$, $\boldsymbol{a}-\boldsymbol{b}-\boldsymbol{c}$, $\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{c}$ *ровно один будет иметь длину, меньшую единицы.*

Доказательство. Пусть $\overrightarrow{AO}=\boldsymbol{a}$, $\overrightarrow{OB}=\boldsymbol{b}$, $\overrightarrow{OC}=\boldsymbol{c}$, $\overrightarrow{OF}=\boldsymbol{b}+\boldsymbol{c}$, $\overrightarrow{OG}=-\boldsymbol{b}+\boldsymbol{c}$, $\overrightarrow{OH}=-\boldsymbol{b}-\boldsymbol{c}$, $\overrightarrow{OI}=\boldsymbol{b}-\boldsymbol{c}$ (рис. 5). Точки F, G, H, I — вершины ромба. Нужно

[1] Как позднее выяснили авторы, формулировка леммы не оригинальна, именно её в качестве ещё одного (уже правильного) решения предложили составители задач АМС.

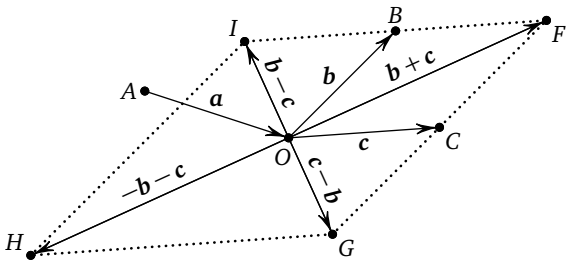


Рис. 5

доказать, что ровно одна из них находится на расстоянии, меньшем единицы, от точки A. Центр ромба — точка O, а сторонам соответствуют векторы $\overrightarrow{FG} = -2\boldsymbol{b}$, $\overrightarrow{GH} = -2\boldsymbol{c}$, $\overrightarrow{HI} = 2\boldsymbol{b}$, $\overrightarrow{IF} = 2\boldsymbol{c}$. Если провести теперь четыре единичных окружности с центрами в F, G, H, I (рис. 6), то первая и вторая, вторая и третья, третья и четвёртая, четвёртая и первая окружности касаются друг друга в серединах сторон ромба (точках C, D, E, B).

Точка A лежит на единичной окружности с центром в точке O. Эта окружность проходит через точки B, C, D, E, которые делят её на четыре дуги. Точка A отлична от точек B, C, D, E (из-за неколлинеарности векторов $\boldsymbol{a}$, $\boldsymbol{b}$ и $\boldsymbol{c}$) и лежит на одной из указанных четырёх дуг. Например, на дуге EB, как на рис. 6. Ясно, что точка A не попадает в единичные круги с центрами в точках F и H (эти круги внешним образом касаются круга с центром в точке I). Не попадает она и в единичный круг с центром

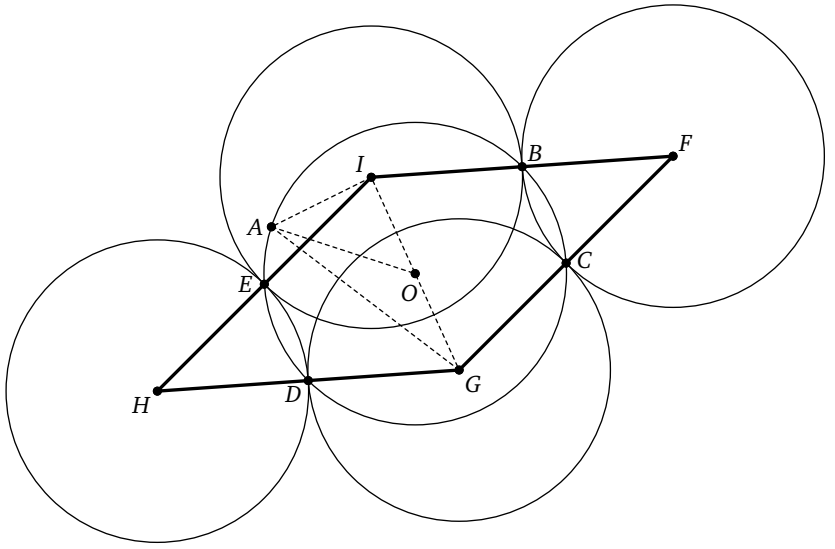


Рис. 6

в точке G. В самом деле, если $AI < 1$, $AG < 1$ и $AO = 1$, то, рассмотрев треугольники AIO и AGO, имеем неравенства для их углов:

$$\angle AIO > \angle AOI, \quad \angle AGO > \angle AOG.$$

Поскольку углы AOI и AOG смежные, сложение этих неравенств даёт

$$\angle AIO + \angle AGO > 180^\circ.$$

Но тогда сумма двух углов в треугольнике AIG больше 180°. Противоречие!

Таким образом, ровно одна из вершин ромба находится на расстоянии, меньшем единицы, от точки A. Это и требовалось доказать. □

Замечание. То, что точка A не попадает в пересечение двух кругов, можно было доказать и таким рассуждением. Пусть какие-то два вектора из указанных четырёх имеют длину меньше 1, например, $|\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{c}| < 1$ и $|\boldsymbol{a}-\boldsymbol{b}+\boldsymbol{c}| < 1$. Тогда

$$2|\boldsymbol{a}| = |(\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{c}) + (\boldsymbol{a}-\boldsymbol{b}+\boldsymbol{c})| \leqslant |\boldsymbol{a}+\boldsymbol{b}-\boldsymbol{c}| + |\boldsymbol{a}-\boldsymbol{b}+\boldsymbol{c}| < 2,$$

откуда $|\boldsymbol{a}| < 1$ — противоречие.

Из леммы следует, что искомая вероятность в задаче про лягушку есть $1/4$. Действительно, пусть в прямоугольной системе координат Oxy направления 1-го, 2-го и 3-го прыжков составляют углы α, β и γ с осью Ox, а $I(\alpha, \beta, \gamma)$ — индикатор интересующего нас события (эта величина равна единице, когда событие происходит, и равна нулю в противном случае). Тогда искомая вероятность P может быть вычислена как

$$P = \frac{1}{(2\pi)^3} \iiint\limits_D I(\alpha, \beta, \gamma)\, d\alpha\, d\beta\, d\gamma,$$

где D — декартов куб отрезка $[0; 2\pi]$. Интеграл от каждой из функций $I(\alpha, \beta+\pi, \gamma)$, $I(\alpha, \beta, \gamma+\pi)$, $I(\alpha, \beta+\pi, \gamma+\pi)$ (аргументы складываются по модулю 2π) даст тот же результат. Поэтому интеграл от суммы индикаторов

$$I(\alpha, \beta, \gamma) + I(\alpha, \beta+\pi, \gamma) + I(\alpha, \beta, \gamma+\pi) + I(\alpha, \beta+\pi, \gamma+\pi)$$

будет в 4 раза больше. Сумма же индикаторов, как показывает лемма, почти всюду равна единице. Отсюда $4P = 1$ и $P = 1/4$.

§ 3. Распределение суммы двух величин, равномерно распределённых на единичной окружности

Вернёмся к задаче нахождения плотности распределения суммы двух случайных величин, каждая из которых равномерно распределена на единичной окружности с центром в начале координат O.

В силу симметрии, плотность распределения $f(x, y)$ в точке $M(x, y)$ зависит только от

$$OM = r = \sqrt{x^2 + y^2}.$$

Пусть α — угол поворота направления второго прыжка относительно первого. Тогда (проверьте это!) $r = 2\cos(\alpha/2)$. Случайная величина $\varphi = \alpha/2$ равномерно распределена на отрезке $[-\pi/2; \pi/2]$. Поэтому для функции распределения случайной величины r имеем

$$F_r(x) = P\{2\cos\varphi < x\} = 2P\left\{\arccos\frac{x}{2} < \varphi \leqslant \frac{\pi}{2}\right\} = \frac{2}{\pi}\cdot\arcsin\frac{x}{2},$$

где $0 \leqslant x \leqslant 2$.

По одномерному закону распределения r двумерную плотность $f(x, y)$ найдём с помощью рассуждения, предложенного проф. Л. Д. Менихесом.

В тождестве

$$\iint\limits_{x^2+y^2<r^2} f(x, y)\,dx\,dy = \frac{2}{\pi}\cdot\arcsin\frac{r}{2}$$

перейдём к полярным координатам (ρ, φ), полагая $f(x, y) = g(\rho)$:

$$\int\limits_0^{2\pi} d\varphi \int\limits_0^r \rho g(\rho)\,d\rho = \frac{2}{\pi}\cdot\arcsin\frac{r}{2}.$$

Отсюда

$$\int\limits_0^r \rho g(\rho)\,d\rho = \frac{1}{\pi^2}\cdot\arcsin\frac{r}{2}.$$

Дифференцирование по r даёт

$$rg(r) = \frac{1}{\pi^2}\cdot\frac{1/2}{\sqrt{1-(r/2)^2}}, \quad \text{т. е.} \quad g(r) = \frac{1}{\pi^2\cdot r\cdot\sqrt{4-r^2}}.$$

Получилось, что плотность распределения бесконечна в центре круга (этого следовало ожидать, поскольку есть бесконечное число способов попасть за два прыжка лягушки из точки O в точку O) и на его границе. Последнее весьма неожиданно!

Теперь (ещё раз!) найдём вероятность попадания в маленький круг

$$P = \int\limits_{-\pi/2}^{\pi/2} d\varphi \int\limits_0^{2\cos\varphi} \rho g(\rho)\,d\rho = \int\limits_{-\pi/2}^{\pi/2} d\varphi \int\limits_0^{2\cos\varphi} \frac{1}{\pi^2}\cdot\frac{1}{\sqrt{4-\rho^2}}\,d\rho = \ldots = \frac{1}{4}.$$

Благодарности

Автор выражает благодарность М. Д. Бронштейну, С. М. Воронину, Э. Ю. Лернеру, Л. Д. Менихесу за продуктивное обсуждение данного сюжета. Из участников «Математического конкурса в ЮУрГУ» с задачей про лягушку справились студенты М. Евгеньев (Челябинск, ЮУрГУ), Е. Кичак (Москва, МИРЭА), К. Чернышёв (Санкт-Петербург, СПбГУ).

Публикация [6] посвящена задачам по теории вероятностей, предлагавшимся в последние годы на различных студенческих олимпиадах.

Список литературы

[1] *Эвнин А. Ю.* 150 красивых задач для будущих математиков. М.: Ленанд, 2018.

[2] *Эвнин А. Ю.* Ещё 150 красивых задач для будущих математиков. М.: Ленанд, 2018.

[3] *Эвнин А. Ю.* Задачи математического конкурса в ЮУрГУ // Математическое образование. 2015. № 4(76). С. 26–52.

[4] *Эвнин А. Ю.* Олимпиада в форме компьютерного теста // Математика в высшем образовании. № 11 (2013). С. 97–102.

[5] *Эвнин А. Ю.* Олимпиада в форме командной игры // Математика в высшем образовании. № 13 (2015). С. 81–94.

[6] *Эвнин А. Ю.*, *Лернер Э. Ю.*, *Игнатов Ю. А.*, *Григорьева И. С.* Задачи по теории вероятностей на студенческих олимпиадах // Математическое образование. 2017. № 4(84). С. 45–60.

Александр Юрьевич Эвнин, Южно-Уральский государственный университет

graph98@yandex.ru

Задачник

(составитель А. Я. Канель-Белов)

Условия задач

В этом разделе вниманию читателей предлагается подборка задач разной степени сложности, в основном трудных. Надеемся, что эти задачи окажутся интересными для читателей «Математического просвещения», в том числе для сильных школьников, интересующихся математикой.

Мы обращаемся с просьбой ко всем читателям, имеющим собственные подборки таких задач, присылать их в редакцию. Мы с удовольствием будем публиковать свежие авторские задачи.

В скобках после условия задачи указывается автор (уточнения со стороны читателей приветствуются).

1. Дан единичный вектор. Разрешается выбрать прямую и заменить вектор его проекцией на эту прямую. С новым вектором можно провести ту же процедуру, и т. д. Прямая каждый раз выбирается заново. Можно ли таким образом развернуть вектор на 180°, потеряв при этом не более 1 % от его длины? (*А. Я. Канель-Белов*)

2. а) Пусть 0 — притягивающая точка непрерывно дифференцируемой функции $f(x)$ (т. е. $0 < |f'(0)| < 1$). Положим $k = f'(0)$. Докажите для всех x_0 из некоторой окрестности нуля существование предела

$$\lim_{n\to\infty} \frac{f^{(n)}(x_0)}{k^n} =: g(x_0),$$

непрерывность функции g и тождество $g(k \cdot g^{(-1)}(x)) = f(x)$.

б) Докажите, что если f бесконечно дифференцируема, то и g тоже бесконечно дифференцируема.

в) Докажите тождество

$$\lim_{n\to\infty} \frac{\sqrt{x}}{2}\cdot\frac{\sqrt{2+\sqrt{x}}}{2}\cdots\frac{\underbrace{\sqrt{2+\sqrt{2+\ldots+\sqrt{x}}}}_{n\text{ корней}}}{2} = \frac{4-x^2}{\sqrt{2\ln\left(\dfrac{x+\sqrt{x^2-4}}{2}\right)}}.$$

(*А. Я. Канель-Белов*)

3. а) Могут ли три прямые разбить круг на 7 частей, равных по площади?

   б) (Открытый вопрос.) Для каждого n определите, могут ли $n+1$ гиперплоскостей разбить n-мерный шар на $2^{n+1}-1$ частей, равных по объёму? (*Н. С. Келлин*)

4. Докажите, что при всех $a_i \geqslant 0$ выполняется *неравенство Карлемана*:

   $$e\cdot(a_1+\ldots+a_n) > a_0+\sqrt{a_0a_1}+\ldots+\sqrt[n+1]{a_0\ldots a_n},$$

   где e — основание натуральных логарифмов.

5. Пусть l — касательная к окружности ABC в точке A; ω — окружность, которая проходит через точки B, C и пересекает прямую l в точках D и E; W, N — середины дуг DE окружности ω. Докажите, что положение ортоцентра треугольника NWA (если он не вырожден) не зависит от положения окружности ω. (*М. Плотников*)

6. Пусть $P_1,\ldots,P_k$ — многочлены от $x_1,\ldots,x_n$, $k<n$, с а) комплексными, б) действительными коэфициентами. Возможно ли равенство многочленов: $P_1^2+\ldots+P_k^2 = x_1^2+\ldots+x_n^2$? (*Фольклор*)

7. Докажите, что при бесконечно многих n набор чисел $1,\ldots,3n$ можно разбить на n групп по три числа так, чтобы одно из чисел каждой группы было равно сумме двух других. (*Фольклор*)

8. а) Какое максимальное число точек можно отметить в единичном трёхмерном кубе так, чтобы все попарные расстояния были строго больше 1? Больше или равны 1?

   б) Аналогичные вопросы для четырёхмерного куба. (*Фольклор*)

9. а) Дана последовательность a_n, $n=1,2,\ldots$ Известно, что при всех $\gamma>1$ выполнено равенство $\lim_{m\to\infty} a_{[\gamma^m]}=0$. Верно ли, что $\lim_{n\to\infty} a_n=0$? (Здесь $[x]$ означает целую часть числа x.) (*Фольклор*)

   б) Назовём число $\beta\in[0,1]$ *хорошим*, если $\beta=\lim_{n\to\infty}\{\alpha^n\}$ при некотором $\alpha>1$ (здесь $\{x\}$ означает дробную часть числа x). Докажите, что множество хороших чисел не более чем счётно. (*А. Я. Канель-Белов*)

10. В k-мерном пространстве отмечены n точек. Разрешается взять прямую, на которой уже лежит не менее t отмеченных точек, отметить любую другую точку на этой прямой и далее повторять эту процедуру. Оказалось, что любую точку пространства можно отметить. При каком наименьшем числе n изначально отмеченных точек это могло случиться? (Число t изначально фиксировано, ответ зависит от t.)
(*И. В. Митрофанов, Ф. В. Петров*)

11. а) Тригонометрический полином P степени n принимает значения из $[-1; 1]$. При этом $P(0) = 1$. Докажите, что $P(x) > \cos nx$ при всех $x \in \left[-\frac{\pi}{2n}; +\frac{\pi}{2n}\right]$. (*Фольклор*)

б) Известно, что $0 < a_0 < a_1 < \ldots < a_n$. Докажите, что тригонометрический многочлен $a_0 + a_1 \cos x + \ldots + a_n \cos nx$ на отрезке $[0; \pi]$ ровно n раз обращается в нуль. (*В. А. Сендеров, А. Я. Канель-Белов*)

12. На конгресс собрались 2019 учёных — химиков и алхимиков. Химик правдив, а алхимик может сказать правду, а может и соврать. Химиков больше. Каждый всё про всех знает. За какое минимальное число вопросов можно установить, кто есть кто? (*Фольклор*)

Исправление условия

К сожалению, в условии задачи 13.8 (выпуск 13, с. 180) допущена опечатка.

Приводим исправленную формулировку.

Задача 13.8. Непрерывная функция f такова, что

$$\int_0^1 x^k f(x)\,dx = 1 \text{ для любого } k \in \{1, \ldots, n-1\}.$$

Докажите, что $\int_0^1 f^2(x)\,dx \geqslant n^2$. (*SEEMOUS* 2008, *Mircea Dan Rus*)

Дополнение к задачнику

Хорошая задача ценна своими связями. Наиболее содержательные задачи открывают новые сюжеты и темы, в рамках которых возникают новые задачи, открываются новые грани. Именно поэтому решение задач

обогащает и оказывается столь полезным, помимо чисто интеллектуальной тренировки[1)].

Публикуем очередные дополнения к задачам.

В выпуске 11 была опубликована

Задача 11.4. d-мерная ладья бьёт по прямым вдоль осей координат.

а) Какое максимальное число ладей можно расставить в d-мерном кубе $n \times \ldots \times n$ так, чтобы они не били друг друга?

Назовём расстановку ладей *полной*, если в ней максимально возможное число ладей.

б) Слоем трёхмерного куба $n \times n \times n$ назовём квадрат $n \times n$, состоящий из клеток с одинаковой третьей координатой. Пусть первые k слоёв заполнены полно (т. е. в них стоят nk ладей). Докажите, что эту расстановку можно продолжить до полной расстановки всего куба.

в) В трёхмерном кубе $n \times n \times n$ расставили ладьи и зафиксировали угловую клетку. Каково максимальное число подкубов с полной расстановкой и той же угловой клеткой? Аналогичный вопрос для d-мерного куба. (*А. Я. Канель*)

В выпуске 23 было опубликовано дополнение к этой задаче:

Задача $11.4'$. Сколько полных расстановок ладей в трёхмерном кубе $4 \times 4 \times 4$? (*А. Ю. Эвнин*)

В этой связи возникают следующие вопросы:

Задача $11.4''$ (на исследование). Какое минимальное число ладей необходимо поставить в кубе $n \times n \times n$, чтобы они били все поля? Что происходит в многомерном кубе? Интересен, во-первых, случай, когда размерность фиксирована, а $n \to \infty$, а во-вторых, когда $n = 2$. (Эта задача относится к теории *кодов*, *исправляющих ошибки*. На практике чаще всего $n = 2$.)

Задача $11.4'''$ (на исследование). В меру стыдливый Вася может соврать не более одного раза. Он задумал натуральное число, меньшее чем n. Петя хочет найти задуманное число или поймать Васю на вранье. Петя может задавать Васе вопросы, требующие ответа «да» или «нет». Каково минимальное необходимое число вопросов? А если Пете надо только определить задуманное число?

[1)] В этом проблема психологических тестов на интеллект: хорошая задача *поучительна*, т. е. её решение, в частности, влияет на возможности решения других задач. Но и наоборот: чем естественнее задача, тем сильнее при её решении сказывается предыдущий опыт. При решении сложных задач требуются идеи по организации процесса решения. Важную роль играет постановка вспомогательных задач. С другой стороны, при решении неестественной задачи исключается эстетика и естественность как фактор поиска.

В выпуске 13 была опубликована

Задача 13.10. *k-параллелепипедом* называется прямоугольный параллелепипед, среди рёбер которого имеется не более k различных. Докажите[2)], что если параллелепипед P можно разрезать на k-параллелепипеды, то длины его рёбер порождают векторное пространство размерности не выше k над $\mathbb{Q}$. (*Л. Радзивиловский, И. Фещенко, Д. Радченко, М. Танцюра*)

Возникает вопрос о разбиениях на кубы:

Задача 13.10′. а) Можно ли разбить плоскость на попарно различные квадраты? (*А. Я. Канель-Белов*)

б) Докажите, что из попарно различных кубов нельзя составить кирпич. (*Фольклор*)

в) (Открытый вопрос.) Можно ли разбить n-мерное пространство ($n>2$) на попарно различные кубы? (Предполагаемый ответ — отрицательный, невозможность для размерности k влечёт невозможность для $n > k$.) (*А. Я. Канель-Белов*)

В выпуске 17 была опубликована

Задача 17.5. На плоскости нарисованы две а) пересекающиеся б) непересекающиеся окружности. Можно ли одной линейкой построить их центры?

Эту тему развивает

Задача 17.5′. На плоскости нарисована кривая G второго порядка и отмечена точка A. Никаких других пометок на плоскости нет.

а) Пусть G — эллипс. Постройте с помощью циркуля и линейки: центр G; вершины G (точки пересечения G с осями симметрии); фокусы; точки касания с G прямых, которые проходят через A и касаются G.

б) Пусть G — парабола. Постройте с помощью циркуля и линейки: вершину; фокус; точки касания с G прямых, которые проходят через A и касаются G; ось симметрии.

в) Пусть G — гипербола. Постройте с помощью циркуля и линейки: центр G; вершины G (точки пересечения с осью симметрии); фокусы;

[2)] Решению этой задачи и близким вопросам были посвящены работы: *Радзивиловский Л. В.*, *Радченко Д. В.*, *Танцюра М.*, *Фещенко И. С.* Разрезание параллелепипеда на бруски // Математическое просвещение. Сер. 3. Вып. 20. М.: МЦНМО, 2016. С. 215–227; а также: *Шаров Ф. А.* Разрезания прямоугольника на прямоугольники с заданными отношениями сторон // Математическое просвещение. Сер. 3. Вып. 20. М.: МЦНМО, 2016. С. 200–214.

асимптоты; точки касания с G прямых, которые проходят через A и касаются G. (*Г. А. Гальперин*)

В выпуске 23 была опубликована

Задача 23.4. Докажите, что середины сторон произвольного описанного четырёхугольника, отличного от квадрата, лежат на эллипсе, касающемся вписанной окружности в двух точках (возможно, мнимых).
(*А. А. Заславский*)

Ниже ещё одна задача на близкую тему:

Задача 23.4′. Четырёхугольник $ABCD$ вписан в окружность с центром O и описан около окружности с центром I. Докажите, что центры вписанных окружностей треугольников OAB, OBC, OCD, ODA лежат на одной окружности. (*Dao Thanh Oai*)

В выпуске 23 была также опубликована

Задача 23.10. а) Полицейский ловит Гангстера в городе, представляющем собой квадрат 10×10, разбитый улицами на квадратные клетки — кварталы. Полицейский видит Гангстера, только если на него натыкается, и оба они ездят только по улицам. Скорость Полицейского в 10 000 раз больше скорости Гангстера. Может ли Полицейский поймать Гангстера за ограниченное время?

б) Тот же вопрос, если потребовать, чтобы путь Полицейского был *конечнозвенной ломаной*. (*А. Я. Канель-Белов*)

В этой связи возникают следующие задачи:

Задача 23.10′ (открытый вопрос). По рёбрам октаэдра бегают паук и муха. Паук видит муху, находясь с ней на одном ребре. Сможет ли он её поймать, если скорость паука в 2,5 раза больше скорости мухи? При каком наименьшем соотношении скоростей он может её поймать?

Задача 23.10′′. (Задача для исследования. Редакции полное решение неизвестно.) Полицейский ловит Гангстера на трёх улицах длиной 1, выходящих из одной точки. Максимальная скорость Полицейского в 2 раза больше максимальной скорости Гангстера, но Полицейский близорук, он видит только часть улицы длиной ε, а Гангстер видит всё вокруг и Полицейского в том числе. Определить, при каких ε Полицейский сможет поймать Гангстера. Рассмотрите тот же вопрос при других соотношениях скоростей. (*По мотивам задачи: Задачник «Кванта», М645, авторы В. Дринфельд, В. Соколов*)

Решения задач из прошлых выпусков

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 этой задачи было опубликовано в «Математическом просвещении», вып. 4, с. 220. К сожалению, это решение неполно (не доказана непустота пересечения интервалов Δ_n). Публикуем другие два решения, принадлежащие ученикам школы № 57 г. Москвы.

Первое решение.

Лемма 1. *Пусть f — бесконечно гладкая функция и в некоторой точке a какая-то её производная (возможно, нулевая) не равна нулю. Тогда найдётся проколотая окрестность точки a, в которой функция f отлична от нуля.*

Доказательство. По условию существует такое k, что $f^{(k)}(a) \neq 0$ и $f^{(m)}(a) = 0$ при всех $m < k$. По формуле Тейлора с остаточным членом в форме Пеано

$$f(x) = \frac{f^{(k)}(a)(x-a)^k}{k!} + o((x-a)^k).$$

Отсюда очевидно, что в некоторой проколотой окрестности точки a знак $f(x)$ совпадает со знаком $f^{(k)}(a)(x-a)^k \neq 0$. □

Лемма 2. *Пусть f — бесконечно гладкая функция. Тогда если x — предельная точка множества нулей функции f, то все производные f в точке x равны нулю.*

Доказательство. Пусть $M = \{x\colon f(x) = 0\}$. Предположим, что x — предельная точка множества M и $f^{(n)}(x) \neq 0$ для некоторого n. Тогда из леммы 1 получаем, что $f(x_0) \neq 0$ в некоторой проколотой окрестности точки x, чего не может быть, так как x — предельная точка множества M. Противоречие. □

Лемма 3. *Пусть f — бесконечно гладкая функция. Известно, что на любом интервале, на котором f не принимает нулевых значений, f — многочлен. Тогда f — многочлен на всей прямой.*

Доказательство. Пусть $M = \{x\colon f(x) = 0\}$. Так как $f(x)$ — непрерывная функция, то множество $\mathbb{R}\backslash M$ открыто, а значит, представимо в виде счётного объединения непересекающихся интервалов $I_k := (a_k; b_k)$[1], на каждом из которых f является многочленом. Пусть существует такое $k \in \mathbb{N}$, что a_k или b_k — предельная точка множества M. Без ограничения общности пусть это a_k. Тогда из леммы 2 получаем, что $\forall\, n \in \mathbb{N}\cup\{0\}\colon f^{(n)}(a_k) = 0$. Так как f — многочлен на I_k, то $f(x) \equiv 0$ на $[a_k; b_k]$, так как иначе $f^{(n)}(x) = \text{const} \neq 0$ при некотором n. Получили противоречие с определением интервалов I_k.

Значит, a_k и b_k при любом k — не предельные точки M, т. е. $a_k = b_p$ и $b_k = a_q$ для некоторых $p, q \in \mathbb{N}$. Иными словами, интервалы I_p, I_k и I_q «примыкают друг к другу». По непрерывности f является многочленом в граничных точках этих интервалов. Выберем одну из этих точек a, и пусть от неё идёт влево последовательность примыкающих интервалов I_{n_k}. Положим $A = \inf a_{n_k}$. Предположим, что $A \neq -\infty$. На каждом из двух соседних интервалов f совпадает с одним и тем же многочленом, который даёт разложение f в ряд Тейлора в граничной точке. Поэтому f является многочленом на $(A; a]$. Так как $f(a_{n_k}) = 0$ при всех k, по непрерывности $f(A) = 0$ и потому A — предельная точка множества M. Рассуждая как выше, приходим к противоречию. Следовательно, $A = -\infty$.

Рассмотрим теперь последовательность примыкающих интервалов I_{m_k}, идущую от той же точки a вправо, и положим $B = \sup b_{n_k}$. Аналогично предыдущему получаем $B = +\infty$. А это означает, что $f(x) \in \mathbb{R}[x]$, что и требовалось доказать. □

Лемма 4. *Пусть f — бесконечно гладкая функция, на некотором отрезке $[A; B]$ не являющаяся многочленом. Тогда найдётся отрезок $[C; D]$ внутри отрезка $[A; B]$, на котором f также не является многочленом и не принимает нулевых значений.*

Доказательство. Пусть $M = \{x\colon f(x) = 0\}$. Как и в лемме 3, можно записать

$$\mathbb{R}\backslash M = \bigsqcup_{k=1}^{\infty} (a_k; b_k).$$

Предположим, что на каждом из этих интервалов f является многочленом. Но тогда из леммы 3 вытекает, что f является многочленом на $[A; B]$ — противоречие. □

[1] Возможно, одно из a_k равно $-\infty$, а одно из b_k равно $+\infty$.

Продолжим решение исходной задачи. Пусть f — бесконечно гладкая функция и при каждом $x \in \mathbb{R}$ некоторая её производная равна нулю. Предположим, что на некотором отрезке $[A; B]$ функция f не является многочленом. По лемме 4 существует отрезок $[x_0; y_0] \subset [A; B]$, на котором f не является многочленом и не обращается в нуль. Тогда и $f^{(1)}|_{[x_0;y_0]} \notin \mathbb{R}[x]$. По лемме 4 существует отрезок $[x_1; y_1] \supset [x_0; y_0]$, на котором $f^{(1)}$ не является многочленом и не обращается в нуль. Аналогично определяем отрезок $[x_n; y_n]$ для всех n.

Таким образом, мы получим последовательность таких вложенных отрезков $[x_0; y_0] \supset [x_1; y_1] \supset [x_2; y_2] \supset \ldots$, что при всех $n \in \mathbb{N} \cup \{0\}$ и при всех $x \in [x_n; y_n]$ имеем $f^{(n)}(x) \neq 0$. У отрезков $[x_n; y_n]$ есть общая точка x, в которой $f^{(n)}(x) \neq 0$ для всех $n \in \mathbb{N} \cup \{0\}$, чего не может быть по условию задачи — противоречие. Значит, $f(x)$ является многочленом на всей оси.

Комментарий. Эта задача хорошо известна. См., например, книгу *Boas R. P. Jr.* A primer of real functions. Washington, D. C.: Mathematical Association of America, 1960. (Carus Mathematical Monograph; Vol. 13). P. 58.

(*И. Украинцев, ученик 11 класса школы № 57 г. Москвы*)

Второе решение. Рассмотрим множество N точек, для которых существует интервал, их содержащий, в котором f — многочлен. Легко доказать, что если это множество совпадает с $\mathbb{R}$, то f — многочлен на всей прямой. Иначе говоря, достаточно показать, что множество $X = \mathbb{R} \setminus N$ пусто. Для этого нам потребуется *теорема Бэра*: *полное метрическое пространство не может быть представлено в виде объединения счётного числа нигде не плотных множеств*.

Лемма 0. *Пусть функция f бесконечно дифференцируема на интервале $(a; c)$. Если f — многочлен на интервалах $(a; b)$ и $(b; c)$, то f — многочлен на $(a; c)$.*

Доказательство. Оба многочлена совпадают с разложением f в ряд Тейлора в точке b. □

Лемма 1. *Все точки множества X — предельные.*

Доказательство. Пусть точка $x \in X$: x — изолированная. Тогда существует такое $\varepsilon > 0$, что $f \notin \mathbb{R}[x]$ на $(x-\varepsilon; x+\varepsilon)$, но каждая точка из $(x-\varepsilon; x)$ содержится в интервале, на котором $f \in \mathbb{R}[x]$. Зафиксируем $y_0 \in (x-\varepsilon; x)$ и соответствующий интервал $(g_0; h_0)$. Положим

$$H = \sup_{h_0 \leqslant h \leqslant x} h\colon f|_{[g_0;h]} \in \mathbb{R}[x].$$

Тогда $H = x$. Действительно, при $h = x - \delta$ $(\delta > 0)$ отрезок $[g; h]$ можно увеличить по лемме 0. Аналогично для правой полуокрестности точки x. Значит, по лемме 0 функция f — многочлен на интервале, содержащем точку x. Но тогда $x \notin X$ — противоречие. □

Лемма 2. *Множество X пусто.*

Доказательство. Множество X замкнуто (так как его дополнение N, очевидно, открыто). Пусть X непусто. Положим $M_n = X \cap S_n$, где $S_n = \{x\colon f^{(n)}(x) = 0\}$. Тогда M_n замкнуто как пересечение замкнутых множеств и $X = \bigcup_n M_n$. В силу теоремы Бэра существует интервал $(c; d)$, пересечение которого с X непусто и содержится в некотором M_n. Легко проверить по индукции, что при $x_0 \in (X \cap [c; d])$ и $k \geqslant n$ обязательно $f^{(k)}(x_0) = 0$. Действительно, при $k = n$ это верно по выбору $(c; d)$, и если это верно при некотором k, то

$$f^{(k+1)}(x_0) = \lim_{x \to x_0} \frac{f^{(k)}(x) - f^{(k)}(x_0)}{x - x_0} = \lim_{x \to x_0,\, x \in X} \frac{f^{(k)}(x) - f^{(k)}(x_0)}{x - x_0} = 0.$$

Если в $(c; d)] \setminus X$ нет отрезков, то

$$(c; d) \in M_n, \quad f|_{(c;d)} \in \mathbb{R}[x]$$

и, следовательно, $X \cap (c; d) = \varnothing$ — противоречие. Значит, существует отрезок $[l_0; k_0] \subset (c; d) \setminus X$. Положим

$$K = \sup_{k_0 \leqslant k \leqslant d} k\colon f|_{[l_0;k]} \in \mathbb{R}[x], \quad L = \inf_{c \leqslant i \leqslant l_0} l\colon f|_{[l;k_0]} \in \mathbb{R}[x].$$

Так как $(c; d) \cap X$ непусто, хотя бы одна из точек K, L не совпадает с концом отрезка $[c; d]$. Пусть это L (без ограничения общности). Если эта точка принадлежит некоторому интервалу, на котором $f(x) \in \mathbb{R}[x]$, то из леммы 0 получаем противоречие с определением L. Значит, $L \in X$. Тогда по доказанному выше $f^{(t)}(L) = 0$ при всех $t \geqslant n$. Значит, степень многочлена, равного $f(x)$ на $[L; K]$, меньше n (так как он совпадает с разложением $f(x)$ в ряд Тейлора в точке L). Аналогично для всех отрезков из $(c; d) \setminus X$. Следовательно, на этом множестве $f^{(n)}(x) = 0$. На $(c; d) \cap X$ это также выполнено. Значит, $f(x) \in \mathbb{R}[x]$ на $(c; d)$ и этот интервал не может пересекаться с X — противоречие. □

Продолжим решение исходной задачи. Из леммы 2 получаем, что любое $x \in \mathbb{R}$ принадлежит интервалу, на котором $f \in \mathbb{R}[x]$. Любой отрезок $[n; n + 1]$, где n целое, можно покрыть конечным числом таких интервалов. Из леммы 0 следует, что $f(x) \in \mathbb{R}[x]$, что и требовалось доказать.

(*А. Соколов, ученик 11 класса школы № 57 г. Москвы*)

6.9′ (выпуск 21). Условие. Прожектор освещает бесконечный угол величиной в 1 градус. Разрешается располагать произвольным образом любое конечное количество прожекторов на плоскости так, чтобы они осветили целиком всю плоскость. Какое наименьшее число прожекторов потребуется, если это а) евклидова плоскость; б) плоскость Лобачевского?
(*Г. А. Гальперин*)

а) Ответ: 360.

Решение. Очевидно, что 360 прожекторов для освещения евклидовой плоскости достаточно. Для этого можно разрезать плоскость на 360 равных углов с общей вершиной и каждый из этих углов осветить своим прожектором.

На самом деле верно гораздо более сильное утверждение. Прожекторы можно разместить так, чтобы они освещали всю плоскость и при этом находились в произвольных наперёд заданных точках. Доказательство этого утверждения можно найти в статье: *Гальперин В.*, *Гальперин Г.* Освещение плоскости прожекторами // Квант. 1981. № 11. С. 28–30.

Рассмотрим теперь произвольные $n < 360$ прожекторов на плоскости и докажем, что они не освещают её полностью. Приведём два способа доказательства этого утверждения.

Первый способ. Сдвинем параллельными переносами все освещаемые прожекторами углы так, чтобы их вершины оказались в одной точке. Поскольку сумма величин этих углов меньше 360°, найдётся луч, выходящий из этой точки и не содержащийся ни в одном из них. При обратном сдвиге любого такого угла он будет пересекаться с найденным лучом не более чем по отрезку. Значит, все прожекторы будут освещать лишь ограниченную область на этом луче. Следовательно, на луче найдутся неосвещённые точки.

Второй способ. Поместим все n прожекторов внутрь круга фиксированного радиуса r и построим концентрический ему круг Ω_R переменного радиуса R. Тогда круг радиуса $R+r$ с центром в точке расположения любого прожектора целиком содержит внутри себя круг Ω_R. Поэтому освещённая прожекторами часть круга Ω_R имеет площадь не более $n \cdot \pi(R+r)^2/360$. При достаточно большом R эта величина будет меньше πR^2 — площади круга Ω_R. (Действительно, обе эти величины выражаются многочленами второй степени от R, причём коэффициент при старшей степени в первом случае меньше.) Значит, этот круг не освещён полностью.

б) Ответ: 2.

Решение. Чтобы разместить два прожектора, освещающих всю плоскость Лобачевского, воспользуемся следующим утверждением (верным в геометрии Лобачевского, но неверным в евклидовой геометрии). Любой

угол содержит внутри себя некоторую прямую. Действительно, каждый угол содержит некоторую дугу абсолюта. Соединив две любые точки этой дуги, получим искомую прямую. Приведём также другое доказательство этого утверждения. Обозначим величину угла через α. Проведём произвольную прямую, а затем найдём точку, для которой угол параллельности относительно данной прямой меньше $\alpha/2$. (Такая точка найдётся в силу теоремы Лежандра, утверждающей, что любой острый угол является углом параллельности относительно данной прямой для некоторой точки.) Далее построим угол величины α с вершиной в этой точке, биссектрисой которого является перпендикуляр из точки на прямую. Из определения угла параллельности следует, что прямая целиком находится внутри угла.

Теперь разместим первый прожектор произвольным образом. Находим прямую, целиком содержащуюся внутри освещаемого им угла. Второй прожектор размещаем симметрично первому относительно этой прямой. Таким образом, каждая из двух полуплоскостей, ограниченных этой прямой, целиком освещается одним из прожекторов, а вместе они освещают всю плоскость.

Очевидно, что одного прожектора для освещения плоскости недостаточно. (*В. О. Бугаенко, Г. А. Гальперин*)

10.7. Условие. От прямоугольника отрезают квадрат, а с оставшимся прямоугольником производят ту же процедуру и далее повторяют её. Является ли последовательность отношений сторон (большей к меньшей) у этих прямоугольников периодической, если одна из сторон исходного прямоугольника равна 1, а другая равна а) $\sqrt{2}$, б) $\sqrt[3]{2}$, в) $\sqrt{2005}$?

а) Ответ: да. Период состоит из двух членов: $\sqrt{2}$ и $\sqrt{2}+1$.

б) Ответ: нет.

Решение. Каждый член последовательности получается из предыдущего либо вычитанием единицы, либо вычитанием единицы и обращением. Рассмотрим подпоследовательность $\{x_n\}$ рассматриваемой последовательности, состоящую из членов, получаемых из предыдущего вторым способом. Эта подпоследовательность может быть задана рекуррентным соотношением $x_{n+1} = 1/(x_n - a_n)$, где a_n — натуральное число. Условие периодичности такой последовательности легко сводится к квадратному уравнению с целыми коэффициентами. Таким образом, если последовательность периодична, то все её члены являются квадратичными иррациональностями, что неверно для $\sqrt[3]{2}$.

в) Ответ: да.

Решение. Верно и утверждение, обратное к доказанному выше: если начальное отношение является квадратичной иррациональностью, то

последовательность периодична. Достаточно доказать, что какой-либо член этой последовательности повторится. Каждый член последовательности может быть получен из начального многократным применением двух операций — вычитания единицы и взятия обратного. Заметим, что если начальный член последовательности является квадратичной иррациональностью, то это же верно для всех её членов. Действительно, если число α — корень многочлена ax^2+bx+c, то $\alpha-1$ является корнем многочлена

$$a(x+1)^2+b(x+1)+c=ax^2+(b+2a)x+(a+b+c),$$

а $1/\alpha$ — корнем многочлена cx^2+bx+a. Будем вместо последовательности чисел рассматривать последовательность квадратных трёхчленов с целыми коэффициентами, корнями которых они являются. Докажем, что с помощью указанных преобразований мы можем получить лишь конечное число квадратных трёхчленов.

Без ограничения общности можно считать, что $a>0$. Непосредственно проверяется, что дискриминант $D=b^2-4ac$ сохраняется при преобразованиях двух описанных типов:

$$(a,b,c)\mapsto(a,b+2a,a+b+c)\quad\text{и}\quad(a,b,c)\mapsto(c,b,a)$$

(или $(a,b,c)\mapsto(-c,-b,-a)$, если $c<0$). Поэтому при фиксированном b фиксировано и произведение $ac=(D-b^2)/4$, а значит, имеется лишь конечное число возможных пар значений коэффициентов a и c. Осталось доказать, что коэффициенты b у всех таких многочленов ограничены.

Преобразование второго типа сохраняет $|b|$. Если $|b|>\sqrt{D}$, то преобразование первого типа уменьшает $|b|$. Докажем это. Заметим, что в этом случае корни рассматриваемого квадратного трёхчлена имеют одинаковый знак, противоположный знаку коэффициента b (это следует из формулы корней квадратного уравнения). В нашем случае один из корней положителен, значит, и второй тоже. Поэтому $b<0$. При преобразовании первого типа к нему прибавляется положительное число $2a$. Если получившийся новый коэффициент $b'=b+2a$ таков, что $|b'|<\sqrt{D}$, то тем самым $|b'|<|b|$. Если же $|b'|>\sqrt{D}$, то в силу изложенного выше наблюдения $b'<0$, и значит, с учётом того, что $b<b'$, опять же $|b'|<|b|$. (Заметим, что поскольку корни многочлена иррациональны, D не является полным квадратом и поэтому случай равенства $|b'|=\sqrt{D}$ невозможен.) Значит, рассматриваемая последовательность обязательно содержит многочлен, для которого $|b|<\sqrt{D}$ (в противном случае абсолютные величины коэффициентов b образовывали бы монотонно убывающую последовательность положительных целых чисел). Назовём часть рассматриваемой последовательности до этого многочлена включительно её *началом*.

Имеется лишь конечное число возможных многочленов, для которых выполняется условие $|b| < \sqrt{D}$. Над каждым из них произведём преобразование первого типа и среди результатов найдём многочлен с максимальным по абсолютной величине коэффициентом b. Этот максимум ограничивает сверху абсолютные величины коэффициентов b всех многочленов последовательности, кроме, быть может, её начала. Утверждение доказано.

Фактически мы доказали, что квадратичные иррациональности — это в точности числа, выражаемые периодическими цепными дробями. Это утверждение называется *теоремой Лагранжа*. Её доказательство, близкое приведённому выше, можно прочитать в книге А. Я. Хинчина «Цепные дроби» (М.: Наука, 1978, с. 62–65). (*В. О. Бугаенко*)

11.4′ (выпуск 23). Условие. Сколько полных расстановок ладей в трёхмерном кубе $4 \times 4 \times 4$? [Расстановка ладей в k-мерном кубе является полной, если ладьи не бьют друг друга и количество ладей максимально возможное.] (*А. Ю. Эвнин*)

Ответ: 576.

Первое решение. Разобьём куб на 16 столбиков, а также на 4 слоя толщиной 1. В каждом столбике должна быть ровно одна ладья. Расположение ладей представим матрицей 4×4, в которой $a_{i,j}$ есть номер слоя, содержащего ладью из данного столбика. Нетрудно видеть, что в каждой строке и каждом столбце этой матрицы по одному разу встречаются числа 1, 2, 3, 4. Значит, эта матрица — латинский квадрат 4-го порядка.

Будем расставлять ладьи по слоям, от нижнего до верхнего. Заполнение k-го слоя означает расстановку в матрице четырёх символов k по одному в каждой строчке и каждом столбце.

Ясно, что после того, как заполнены три слоя, последний заполняется однозначно (ставим ладьи в свободные от них столбцы, или, что то же самое, записываем четвёрки в свободные клетки матрицы). Первый слой можно задать перестановкой (j_1, j_2, j_3, j_4) чисел от 1 до 4, где $a_{i,j_i} = 1$ для всех i. Поэтому первый слой заполняется $4! = 24$ способами.

Второй слой задаётся перестановкой π чисел от 1 до 4, которая по отношению к перестановке (j_1, j_2, j_3, j_4) является беспорядком (на 1-м месте не j_1, на 2-м месте не j_2, …, на 4-м месте не j_4). Число таких беспорядков равно 9. Проблема в том, что количество способов заполнить третий слой зависит от того, как был заполнен второй. Поскольку неподвижных элементов нет, здесь всего два варианта: либо π представляет собой цикл длиной 4 (таких перестановок $3! = 6$), либо произведение двух циклов длиной 2 (таких перестановок 3: для фиксированного элемента выби-

раем пару, что можно сделать тремя способами, оставшиеся элементы образуют вторую пару). Вот примеры заполнения слоёв обоих видов:

$$\begin{pmatrix} 1 & 2 & & \\ & 1 & 2 & \\ & & 1 & 2 \\ 2 & & & 1 \end{pmatrix}; \quad \begin{pmatrix} 1 & 2 & & \\ 2 & 1 & & \\ & & 1 & 2 \\ & & 2 & 1 \end{pmatrix}.$$

Проверьте, что в первом случае третий слой заполняется двумя способами, а во втором — четырьмя.

Поэтому ответ к задаче вычисляется так: $24 \cdot (6 \cdot 2 + 3 \cdot 4) = 576$.

Замечание. В настоящее время (на начало 2019 г.) вычислено количество латинских квадратов лишь до 11 порядка. Так что если в условии задачи заменить числа 4 и 16 на k и k^2 соответственно, то при $k > 11$ науке на данный момент ответ неизвестен! (*А. Ю. Эвнин*)

Второе решение. Сведём задачу к эквивалентной: «Сколько существует способов расставить по четыре элемента каждого из четырёх типов на доске 4×4 так, чтобы в каждой строке и каждом столбце был ровно один элемент каждого типа?»

Зафиксируем порядок элементов в первой строке, а затем в оставшихся клетках первого столбца. Теперь выберем определённый элемент и проверим, на каких позициях в оставшемся квадрате 3×3 он может стоять. Непосредственно видим, что ровно в четырёх из шести расстановок позиции остальных элементов определяются однозначно, а в остальных двух возникают противоречия.

Теперь можно менять порядок элементов первой строки, переставляя столбцы 4! способами, и в каждом случае можно менять порядок оставшихся элементов первого столбца, переставляя 3! способами строки 2–4. Окончательно получим $4 \cdot 3! \cdot 4! = 24 \cdot 24 = 576$ расстановок. (*Д. Козырев*)

14.6′ (выпуск 23). Условие. Многочлен $P(x, y)$ от двух переменных принимает только положительные значения. Может ли он принимать все положительные значения? (*Фольклор*)

Ответ: может.

Решение. Пусть $P(x, y) = x^2 + (xy - 1)^2$. Тогда $P(x, y) > 0$, так как оба слагаемых неотрицательны и не могут быть равны нулю одновременно. С другой стороны, для любого положительного числа a возьмём $x = \sqrt{a}$, $y = 1/\sqrt{a}$, и тогда $P(x, y) = a$. (*В. О. Бугаенко*)

14.11. Условие. Известно, что если зафиксировать одну переменную, то функция $f(x, y)$ по другой будет многочленом. Верно ли, что она будет сама многочленом от двух переменных? (*Б. П. Панеях*)

Ответ: да, верно.

Решение.

Лемма 1. *Если последовательность многочленов $\{P_n(x)\}$, степени которых ограничены в совокупности, поточечно сходится на прямой к некоторой функции $f(x)$, то эта функция — многочлен.*

Доказательство. Пусть $\max(\deg P_n) = N-1$. Применив интерполяционную формулу Лагранжа к точкам $1, 2, \ldots, N$, получим для любого x и для всех натуральных n:

$$P_n(x) = \sum_{i=1}^{N} P_n(i) R_i(x), \quad \text{где} \quad R_i(x) = \prod_{j=1,\, j\neq i}^{N} \frac{x-j}{i-j}.$$

Тогда

$$f(x) = \lim_{n\to\infty} P_n(x) = \sum_{i=1}^{N} R_i(x) \lim_{n\to\infty} P_n(i) = \sum_{i=1}^{N} f(i) R_i(x),$$

т. е. $f(x) \in \mathbb{R}[x]$, что и требовалось доказать. □

Лемма 2. *Пусть все числа $x_1, x_2, \ldots, x_n$ различны, причём*

$$\begin{cases} a_0 + x_1 a_1 + x_1^2 a_2 + \ldots + x_1^n a_n = y_1, \\ a_0 + x_2 a_1 + x_2^2 a_2 + \ldots + x_2^n a_n = y_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0 + x_{n+1} a_1 + x_{n+1}^2 a_2 + \ldots + x_{n+1}^n a_n = y_{n+1}. \end{cases}$$

Тогда эта система уравнений относительно $\{a_i\}$ имеет единственное решение, причём a_i представимы в виде линейных комбинаций величин $y_1, y_2, \ldots, y_n$, коэффициенты которых — рациональные функции от x_i.

Доказательство. Рассмотрим (единственный) интерполяционный многочлен Лагранжа, график которого проходит через точки $(x_1; y_1)$, $(x_2; y_2)$, $\ldots$, (x_n, y_n). Согласно системе уравнений из условия, его коэффициенты равны $a_0, a_1, \ldots, a_n$. Из явной формулы интерполяционного многочлена Лагранжа следует утверждение леммы.

Можно также воспользоваться тем, что определитель системы — это определитель Вандермонда, который отличен от нуля, поэтому решение системы единственно. □

Пусть функция $g\colon \mathbb{R}^2 \to \mathbb{R}$ такова, что $g(x_0, y) \in \mathbb{R}[y]$ при любом $x_0 \in \mathbb{R}$ и $g(x, y_0) \in \mathbb{R}[x]$ при любом y_0, принадлежащем некоторому отрезку $[a; b]$. Положим $F(x, y) := \partial g(x, y)/\partial y$. Очевидно, $F(x_0, y) \in \mathbb{R}[y]$ при всех $x_0 \in \mathbb{R}$.

Лемма 3. *Существует такой отрезок* $[c;d]\subset[a;b]$, *что* $F(x,y_0)\in\mathbb{R}[x]$ *при любом* $y_0\in[c;d]$.

Доказательство. Положим

$$M_k=\{y_0\mid y_0\in[a;b],\deg_x(g(x,y_0))\leqslant k\}.$$

Тогда $\bigcup_{k=1}^{\infty}M_k=[a;b]$. Применим теорему Бэра: полное метрическое пространство не может быть представлено в виде счётного объединения нигде не плотных множеств. Получаем, что некоторое M_k плотно на некотором отрезке $[c;d]\subset[a;b]$. Пусть $y_0\in[c;d]$. При любом $x\in\mathbb{R}$ имеем

$$F(x,y_0)=\frac{\partial g(x,y)}{\partial y}\bigg|_{y=y_0}=\lim_{z\to y_0}\frac{g(x,z)-g(x,y_0)}{z-y_0}.$$

Так как M_k всюду плотно на $[c;d]$, то существует последовательность его элементов $\{z_m\}$, сходящаяся к y_0. Выражение

$$\frac{g(x,z_m)-g(x,y_0)}{z_m-y_0}$$

при любом m обозначает многочлен. По построению M_k множество степеней этих многочленов ограниченно. Как мы показали выше, они поточечно сходятся к $F(x,y_0)$. По лемме 1 получаем $F(x,y_0)\in\mathbb{R}[x]$, что и требовалось доказать. □

Продолжим решение исходной задачи. Пусть $f(x,y)$ удовлетворяет условию задачи. Положим $L_K=\{x_0\mid\deg_y(f(x_0,y))\leqslant K\}$. Так как L_K в совокупности покрывают $\mathbb{R}$, некоторое L_K бесконечно. Пусть $a_0=0$, $b_0=1$, $W_k(x,y)=\partial^k f(x;y)/\partial y^k$. Так как $f(x_0,y)\in\mathbb{R}[y]$, имеем $W_k(x_0,y)\in\mathbb{R}[y]$ при всех k. По лемме 3 имеем $W_1(x,y_0)\in\mathbb{R}[x]$ для всех y_0 из некоторого отрезка $[a_1;b_1]\subset[a_0;b_0]$. Аналогично $W_2(x,y_0)\in\mathbb{R}[x]$ для всех y_0 из некоторого отрезка $[a_2;b_2]\subset[a_1;b_1]$, и т. д.

Возьмём произвольное $y_0\in[a_{K+1};b_{K+1}]$. По построению $W_{K+1}(x,y_0)\in\in\mathbb{R}[x]$. С другой стороны, для любого $x_0\in L_K$ имеем $\deg(f(x_0,y))\leqslant K$, откуда $W_{K+1}(x_0,y)=0$ при любом y. Так как L_K бесконечно, найдётся бесконечно много таких x_0, что $W_{K+1}(x_0,y_0)=0$. Значит, многочлен $W_{K+1}(x,y_0)$ тождественно равен нулю. Но y_0 — произвольная точка из $[a_{K+1};b_{K+1}]$. Поэтому для любого $x\in\mathbb{R}$ многочлен $W_{K+1}(x,y)$ тождественно равен нулю на $[a_{K+1};b_{K+1}]$ и, следовательно, тождественно равен нулю на $\mathbb{R}$. Таким образом, для любого $x\in\mathbb{R}$ имеем $\partial^{K+1}f(x;y)/\partial y^{K+1}\equiv 0$ на $\mathbb{R}$ и потому степень многочлена $f(x,y)$ по y не превосходит K. Тогда можно записать

$$f(x,y)=\sum_{i=0}^{K}a_i(y)x^i,$$

и выполнена система уравнений

$$\begin{cases} a_0(y)+a_1(y)+a_2(y)+\ldots+a_K(y)=f(1,y)\in\mathbb{R}[y], \\ a_0(y)+2a_1(y)+4a_2(y)+\ldots+2^K a_K(y)=f(2,y)\in\mathbb{R}[y], \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0(y)+(K+1)a_1(y)+(K+1)^2a_2(y)+\ldots+(K+1)^K a_K(y)= \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad =f(K+1,y)\in\mathbb{R}[y]. \end{cases}$$

По лемме 2 каждое $a_i(y)$ представимо в виде линейной комбинации функций $f(1,y)$, $f(2,y)$, …, $f(K+1,y)\in\mathbb{R}[y]$. Значит, $a_i(y)\in\mathbb{R}[y]$ для всех i. А так как

$$f(x,y)=\sum_{i=0}^{N}a_i(y)x^i,$$

получаем, что $f(x,y)\in\mathbb{R}[x,y]$, что и требовалось доказать.

Комментарий. Для функций, совпадающих с многочленами в рациональных точках, утверждение задачи, вообще говоря, неверно:

Утверждение. *Существует такая аналитическая функция* $g\colon\mathbb{R}^2\to\mathbb{R}$, *что* $g(x_0,y)\in\mathbb{R}[y]$, $g(x,y_0)\in\mathbb{R}[x]$ *при всех* $x_0,y_0\in\mathbb{Q}$, *но* $g(x,y)\notin\mathbb{R}[x,y]$.

Доказательство. Пусть $z_1,z_2,\ldots$ — нумерация всех рациональных чисел. Для каждого $n\in\mathbb{N}$ существует такое $k_n\in\mathbb{N}$, что $|z_j|/k_n\leqslant 1$ при всех $j\leqslant n$. Рассмотрим функцию

$$g(x,y)=\sum_{n=1}^{\infty}\prod_{j=1}^{n}\frac{(x-z_j)(y-z_j)}{n!k_n^2}.$$

Покажем, что данный ряд сходится абсолютно на всей плоскости. Действительно, при всех $n\in\mathbb{N}$, $x,y\in\mathbb{R}$

$$\prod_{j=1}^{n}\frac{|(x-z_j)(y-z_j)|}{n!k_n^2}\leqslant\prod_{j=1}^{n}\frac{|xy|+|xz_j|+|z_jy|+|z_j^2|}{n!k_n^2}\leqslant$$
$$\leqslant\prod_{j=1}^{n}\frac{|xy|+|x|+|y|+|1|}{n!}=\frac{(|xy|+|x|+|y|+|1|)^n}{(n!)^n}.$$

Значит,

$$\left|\sum_{n=1}^{\infty}\prod_{j=1}^{n}\frac{(x-z_j)(y-z_j)}{n!k_n^2}\right|\leqslant\sum_{n=1}^{\infty}\frac{(|xy|+|x|+|y|+|1|)^n}{(n!)^n}.$$

Ряд в правой части очевидно сходится. Следовательно, функция $g(x,y)$ определена корректно и является аналитической.

Для любых $i, n \in \mathbb{N}$, $n \geqslant i$ имеем

$$\prod_{j=1}^{n} \frac{|(z_i - z_j)(y - z_j)|}{n! k_n^2} = 0,$$

поэтому $g(z_i, y) \in \mathbb{R}[y]$. Аналогично $g(x, z_i) \in \mathbb{R}[x]$. С другой стороны, для любого натурального N имеем

$$\deg_x(g(x, z_{N+2})) = N + 1 > N.$$

Следовательно, $g \notin \mathbb{R}[x, y]$. □

(*И. Украинцев, ученик 11 класса школы № 57 г. Москвы*)

15.9. Условие. а) Дана 2×2-матрица A с вещественными коэффициентами. Докажите, что её можно представить как сумму квадратов двух матриц второго порядка с вещественными коэффициентами.

(*SEEMOUS 2010*)

б)* Можно ли матрицу размера $n \times n$ с вещественными коэффициентами представить в виде суммы квадратов нескольких матриц размера $n \times n$ с вещественными коэффициентами? Если «да», то каково минимальное число квадратов? (*О. Ливнэ Бар-Он, Ш. Кармиели*)

а) Решение. Любую вещественную 2×2-матрицу можно представить в виде

$$A = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} + \begin{pmatrix} k & m \\ 0 & l \end{pmatrix},$$

где k и l положительны. Матрицы первого вида находятся в естественном соответствии с комплексными числами, сохраняющем умножение. Из комплексного числа всегда извлекается квадратный корень, поэтому он извлекается из первой матрицы. Квадратный корень из второй матрицы равен

$$\begin{pmatrix} \sqrt{k} & \dfrac{m}{\sqrt{k} + \sqrt{l}} \\ 0 & \sqrt{l} \end{pmatrix}.$$

(*О. Ливнэ Бар-Он, Ш. Кармиели*)

б) Ответ: при $n = 1$ указанное представление не всегда возможно (отрицательное число непредставимо как сумма квадратов).

Для остальных нечётных n: необходимы три квадрата.

Для чётных n: необходимы два квадрата.

Решение. Пусть n чётно.

Лемма. *В некоторой ε-окрестности единичной матрицы квадратный корень извлекается.*

Доказательство. Применим формулу бинома Ньютона:

$$(1+x)^{1/2} = f_{1/2}(x) = \sum_{n=0}^{\infty} \binom{1/2}{n} x^n.$$

Радиус сходимости этого ряда равен 1, поэтому для матрицы с нормой, меньшей 1, ряд сходится. Непосредственно проверяется, что $(f_{1/2}(x))^2 = 1 + x$, и аналогичное верно для матриц, когда ряд сходится (поскольку степени матрицы коммутируют). Лемма доказана. □

Обозначим через m максимум абсолютных величин элементов матрицы A. Пусть M таково, что $1/M^2 < \varepsilon/m$. Положим $A = D + (-M^2E)$, где E — единичная матрица. Тогда $D = M^2(E + A/M^2)$. Но $E + A/M^2$ находится в ε-окрестности единичной матрицы, и согласно лемме из неё извлекается квадратный корень, а тогда он извлекается и из D.

Второе слагаемое, $-M^2I$, также является квадратом, так как в чётных размерностях квадратом является $-E$ (при $n = 2$ это поворот на 90°, а в высших размерностях используем разбиение на двумерные блоки). Таким образом, любую 2×2-матрицу можно разложить в сумму двух квадратов. Квадратом является не каждая такая матрица (матрицы с отрицательным определителем квадратами не являются).

Пусть теперь n нечётно. Докажем, что $-E$ не является суммой двух квадратов, так что иногда нужно не меньше трёх, и построим разложение произвольной матрицы на три квадрата.

Разложение строится аналогично чётному случаю. Пусть V — диагональная матрица, причём в правом нижнем углу стоит 1, а остальные диагональные элементы равны -9. Далее, пусть U — диагональная матрица, в левом верхнем углу которой стоит 1, а остальные диагональные элементы равны -9. У этих матриц есть квадратный корень: возьмём диагональную матрицу размерности на 1 меньше, у которой на диагонали стоит -1, извлечём из неё квадратный корень, умножим на 3 и добавим единичную 1×1-матрицу.

Любую $n \times n$-матрицу A можно представить в виде $A = MV + MU + W$ с любым достаточно большим положительным M. Но если M достаточно велико, то W/M близко к диагональной матрице, у которой на концах диагонали стоит 8, а остальные диагональные элементы равны 18. Согласно лемме из неё извлекается квадратный корень. Искомое разложение построено.

Осталось доказать, что $-E$ не является суммой двух квадратов. Пусть $E = A^2 + B^2$, где A, B — вещественные матрицы. Выберем базис над полем комплексных чисел, в котором матрица A верхнетреугольная. Тогда

A^2 и $B^2 = -E - A^2$ тоже верхнетреугольные. На главной диагонали этих матриц стоят квадраты собственных значений матриц A и B. Получаем систему уравнений: $a_k^2 + b_k^2 = -1$.

Так как многочлен нечётной степени имеет нечётное количество вещественных корней (с учётом кратности), то среди a_k нечётное количество вещественных, а остальные разбиваются на пары комплексно сопряжённых. То же верно для b_k. Но если a_k вещественное, то $a_k^2 \geqslant 0$, тогда $b_k^2 \leqslant -1$, т. е. b_k не является вещественным. Оно комплексно сопряжено некоторому b_j, тогда $a_j^2 \geqslant 0$ и a_j вещественное. Таким образом, вещественные собственные значения матрицы A разбиваются на пары, что невозможно, так как их количество нечётно. Получено противоречие, что и требовалось. (*О. Ливнэ Бар-Он, Ш. Кармиели*)

16.5. Условие. Двумерная фигура в четырёхмерном пространстве имеет площадь S. Её проекция на плоскость первых двух координатных осей имеет площадь S_1, а проекция на плоскость последних двух осей имеет площадь S_2. Докажите, что $S \geqslant S_1 + S_2$. (*Фольклор*)

Решение. Измеримая фигура на двумерной поверхности может быть аппроксимирована прямоугольниками, и то же верно для её проекций на координатные плоскости. Поэтому достаточно доказать неравенство для прямоугольника. Пусть он задан векторами $u = (u_1, u_2, u_3, u_4)$ и $v = (v_1, v_2, v_3, v_4)$. Тогда

$$S = |u| \cdot |v|, \quad \pm S_1 = u_1 v_2 - u_2 v_1, \quad \pm S_2 = u_3 v_4 - u_4 v_3.$$

Положим $u^* = (\pm u_2, \pm u_1, \pm u_4, \pm u_3)$, где знаки выбраны таким образом, что $S_1 + S_2$ равно скалярному произведению $\langle u^*, v \rangle$. Поскольку

$$\langle u^*, v \rangle \leqslant |u^*| \cdot |v| = |u| \cdot |v| = S,$$

получаем нужное неравенство. (*Л. Радзивиловский*)

19.6. Условие. Пусть $\alpha \notin \mathbb{Q}$ — иррациональное число, β — произвольное число из интервала $(0; 1)$, а Q_M — минимум дробной части $N \cdot \alpha$, где $N < M$ — целое. Аналогично R_M есть минимум дробной части $\beta - N \cdot \alpha$. Докажите, что $Q_M > R_M$ при бесконечно многих M. (*С. В. Конягин*)

Первое решение. Нам нужно доказать, что для любого натурального числа N найдётся такой номер $n > N$, что $Q_n > R_n$. Если $R_N = 0$, то $R_n = 0 < Q_n$ при всех $n > N$. Пусть $R_N > 0$. В силу иррациональности α любой интервал $(a; b) \subset (0; 1)$ содержит бесконечно много чисел $\{n\alpha\}$. В частности, существуют такие натуральные числа n, что

$$\max(0, \beta - R_N) < \{n\alpha\} < \beta, \quad \text{т. е.} \quad \{\beta - n\alpha\} < \min(\beta, R_N).$$

Среди таких чисел n выберем минимальное n_0. Число n_0 удовлетворяет условиям

$$\{\beta - n_0\alpha\} < \min(\beta, R_N) \leqslant \{\beta - n\alpha\}$$

при любом $n = 0, 1, \ldots, n_0 - 1$. Заметим, что $n_0 > N$, так как в противном случае по определению R_N выполнялось бы неравенство $R_N \leqslant \{\beta - n_0\alpha\}$.

Если $R_{n_0} < Q_{n_0}$, то искомое $n = n_0$ найдено. Допустим, что $R_{n_0} \geqslant Q_{n_0} = \{m\alpha\}$, $1 \leqslant m \leqslant n_0$. Тогда выполняются соотношения

$$R_{n_0+m} \leqslant \{\beta - (n_0 + m)\alpha\} = R_{n_0} - Q_{n_0}, \qquad (*)$$

$$Q_{n_0+m} = Q_{n_0} = \{m\alpha\}. \qquad (**)$$

Действительно, неравенство $(*)$ вытекает из того, что

$$\{\beta - (n_0 + m)\alpha\} = \{\{\beta - n_0\alpha\} - \{m\alpha\}\} = \{R_{n_0} - Q_{n_0}\} = R_{n_0} - Q_{n_0}.$$

Допустим, что равенство $(**)$ не выполнено, т. е. $\{k\alpha\} < Q_{n_0}$ при некотором $k = \{n_0 + 1, \ldots, n_0 + m\}$. Тогда $0 \leqslant n_0 + m - k < n_0$ и

$$\{\beta - (n_0 + m - k)\alpha\} \leqslant \{\beta - (n_0 + m)\alpha\} + \{k\alpha\} = R_{n_0} - Q_{n_0} + \{k\alpha\} < R_{n_0},$$

но это противоречит выбору n_0. Положим $n_1 = n_0 + m$. Если $R_{n_1} < Q_{n_1}$, то остаётся взять $n = n_1$. Если же $R_{n_1} \geqslant Q_{n_1}$, то мы можем подставить в $(*)$ и $(**)$ n_1 вместо n_0. Далее положим $n_2 = n_1 + m = n_0 + 2m$ и т. д.

При некотором k имеем $R_{n_k} < Q_{n_k}$, так как если при всех k будет $R_{n_k} \geqslant Q_{n_k}$, то, подставляя в $(*)$ и $(**)$ вместо n_0 последовательно $n_1, n_2, \ldots$, мы получим

$$Q_{n_0} = Q_{n_0+m} = Q_{n_0+2m} = \ldots,$$

$$R_{n_0+km} \leqslant R_{n_0} - kQ_{n_0}. \qquad (***)$$

Но при больших k правая часть $(***)$ меньше нуля. Это противоречие завершает доказательство.

(*По кн.: В. А. Садовничий, А. А. Григорян, С. В. Конягин. Задачи студенческих математических олимпиад. М.: МГУ, 1987. С. 288*)

Второе решение. Пусть по окружности единичной длины с отмеченной точкой β, начиная из нулевой точки движется Кентавр, длина его прыжка α. Нужно показать, что наилучшее приближение к нулевой точке после n прыжков оказывается лучше, чем к точке β, при бесконечно многих n. Если $\beta \geqslant \alpha$, то можно заменить β на $\beta' = \beta - \alpha$, при этом наилучшее приближение не ухудшится, а количество прыжков уменьшится на 1. Кроме того, можно считать, что $\alpha < 1/2$ (иначе, изменив направление прыжков, перейдём к величине $\alpha' = 1 - \alpha < 1/2$.

Итак, $0 < \beta < \alpha < 1/2$ (случаи равенств очевидны). Рассмотрим дугу $[\alpha; 2\alpha]$. Вырежем её и заклеим. При этом мы не потеряем наилучшие приближения ни к 0, ни к точке β, а количество шагов уменьшится. Мы получили движение по окружности длины $1-\alpha$ с шагом α.

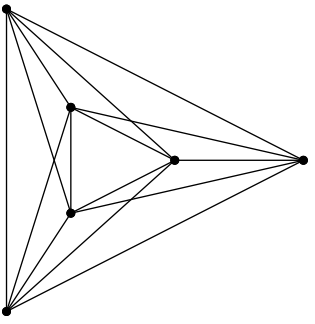
Поскольку одно n, требуемое в условии задачи, очевидным образом существует, дело завершает индукция.

Комментарий. Описанная процедура «режем-клеим» называется *индукцией Рози*. С её помощью проще всего доказать (и понять), что цепные дроби задают наилучшие приближения. Иными словами, пусть дан единичный отрезок и отрезок длины $\alpha \leqslant 1$. Заменяем 1 на $\alpha_1 = 1 - n\alpha$, где $n = [\alpha^{-1}]$ (т. е. откладываем отрезки длины α, пока возможно), от пары $(1, \alpha)$ переходим к новой паре отрезков (α, α_1) и продолжаем ту же процедуру. Утверждается, что α_i дают наилучшие приближения целых чисел числами, кратными α. (См. также задачу 10.7 и её решение, опубликованное в данном выпуске (с. 186), где дано наглядное представление алгоритма Евклида для отрезков, отвечающее процедуре построения цепной дроби.) (*А. Я. Канель-Белов*)

22.3. Условие. а) Для какого максимального N можно расположить N точек на плоскости, соединить их попарно отрезками и раскрасить отрезки в синий и красный цвет так, чтобы отрезки одного цвета не имели внутренних точек пересечения, а каждый отрезок имел не более одной точки пересечения с отрезком другого цвета? (*Л. Радзивиловский*)

Ответ: $N = 6$.

Решение. Задача равносильна следующей: при каких N существует полный граф на плоскости, у которого каждое ребро пересечено не более одного раза и никакие три ребра не пересекаются в общей внутренней точке. Действительно, в таком графе на каждом «перекрёстке» можно покрасить образующие его рёбра в разные цвета. Покрасив остальные рёбра произвольно в красный и синий цвет, получим граф, удовлетворяющий условию исходной задачи.



На рисунке показан пример для новой задачи с $N = 6$. Удаляя из этого графа вершины и примыкающие к ним рёбра, можно получить примеры для меньших N.

Предположим, что имеется полный граф на 7 вершинах, удовлетворяющий условию новой задачи. У него 21 ребро с X точками пересечения. Пусть O — одна из них, принадлежащая рёбрам AC и BD. Рассмотрим четырёхугольник

$ABCD$. Его стороны не пересечены. Действительно, пусть, например, какое-то ребро входит в треугольник ABO, пересекая сторону AB. Тогда оно должно выйти из треугольника через AB, AO или BO и снова пересечь ребро, которое уже пересечено (случай прохождения через вершину треугольника можно исключить малым шевелением).

Таким образом, в графе с минимальным числом пересечений вокруг каждого пересечения расположено четыре непересечённых ребра. Каждое непересечённое ребро при этом засчитывается не более двух раз, поэтому количество непересечённых рёбер не меньше $2X$. Количество пересечённых рёбер равно $2X$, поэтому $4X < 21$, откуда $X \leqslant 5$.

Рассмотрим граф, вершинами которого служат вершины исходного графа и точки пересечения, а рёбрами — непересечённые рёбра исходного графа и половинки пересечённых рёбер. Этот граф, в отличие от исходного, планарен. У него $V = 7 + X$ вершин, $E = 21 + 2X$ рёбер и $F \geqslant 2E/3$ граней. По формуле Эйлера

$$2 = F - E + V \geqslant V - \frac{E}{3} = 7 + X - \left(7 + \frac{2X}{3}\right) = \frac{X}{3},$$

откуда $6 \geqslant X$. Но выше доказано, что $X \leqslant 5$, — противоречие.

Поскольку невозможно построить искомый граф для $N = 7$, это невозможно и для $N > 7$. (*Л. Радзивиловский*)

23.7. Условие. Даны многочлены $P(x)$, $Q(x)$ с неотрицательными коэффициентами и старшим коэффициентом 1. Докажите, что все их коэффициенты равны 0 или 1, если $P(x)Q(x) = 1 + x + x^2 + \ldots + x^n$.

(*Б. И. Каневский*, *В. А. Сендеров*)

Решение. Коэффициенты многочленов $P(x)$ и $Q(x)$ вещественны, поэтому для каждого корня каждого из этих многочленов найдётся комплексно сопряжённый. Но все эти корни равны 1 по абсолютной величине, поэтому можно переформулировать утверждение задачи следующим образом: если z — корень одного из многочленов $P(x)$ и $Q(x)$, то $1/z$ — корень того же многочлена, имеющий ту же кратность. Значит, если

$$P(x) = a_m x^m + a_{m-1} x^{m-1} + \ldots + a_2 x^2 + a_1 x + a_0,$$

то последовательность $\{a_j\}$ симметрична, $a_j = a_{m-j}$; аналогично: если

$$Q(x) = b_n x^k + \ldots + b_1 x + b_0,$$

то $b_j = b_{k-j}$. Разумеется, здесь $m + k = n$.

Мы хотим доказать, что все коэффициенты многочленов $P(x)$ и $Q(x)$ равны 0 или 1. Предположим противное. Эти коэффициенты не могут быть больше единицы: если $a_j > 1$, то коэффициент при x^{k+j} многочлена

$P \cdot Q$ равен a_j плюс неотрицательное число, но он должен быть равен 1. Отметим, что $a_0 = 1 = b_0$, поскольку эти коэффициенты не превосходят 1, неотрицательны и их произведение равно 1.

Выберем наименьшее t, при котором x^t в одном из многочленов имеет коэффициент, отличный от 0 и 1; без ограничения общности можно считать, что это происходит в $P(x)$. Тогда все коэффициенты при меньших степенях в P и Q равны 0 или 1; это $a_0, a_1, \ldots, a_{t-1}$ и $b_0, b_1, \ldots, b_{t-1}$.

Рассмотрим коэффициент при x^t в $P \cdot Q$ (равный 1, как и все коэффициенты этого произведения); он равен $1 \cdot a_t + b_t \cdot 1 +$ произведения нулей и единиц. Так как $a_t > 0$, получаем $b_t = 1 - a_t$, откуда $0 < b_t < 1$. По симметрии $b_{k-t} = b_t$, откуда $0 < b_{k-t} < 1$.

Теперь рассмотрим коэффициент при x^k в $P \cdot Q$, также равный 1. Его слагаемыми являются $b_k a_0 = 1$, строго положительное $b_{k-t} a_t$ и неотрицательные слагаемые. Противоречие. (*Л. Радзивиловский*)

23.9. Условие. Пусть α, β — положительные числа. Рассмотрим такую симметрическую матрицу (a_{ij}), что

$$a_{ij} = \frac{1}{i+j+\alpha} \cdot \frac{1}{i+j+\beta}.$$

Докажите, что эта матрица положительно определена. (*А. А. Логунов*)

Решение. Рассмотрим меру $\frac{dx}{x^{1-\alpha}} \cdot \frac{dy}{y^{1-\beta}}$ на единичном квадрате $[0;1]^2$. При положительных α и β

$$\int_0^1 \int_0^1 \frac{dx}{x^{1-\alpha}} \cdot \frac{dy}{y^{1-\beta}} = \alpha x^\alpha \Big|_0^1 \cdot \beta y^\beta \Big|_0^1 = \alpha\beta.$$

Поэтому любая функция, непрерывная на единичном квадрате и, следовательно, ограниченная, будет интегрируемой по этой мере. Значит, для многочленов можно определить скалярное произведение

$$\langle f, g \rangle = \int_0^1 \int_0^1 f(x,y) \cdot g(x,y) \frac{dx}{x^{1-\alpha}} \cdot \frac{dy}{y^{1-\beta}}.$$

Оно положительно определено, поскольку ненулевой многочлен не может быть тождественным нулём на квадрате. Пусть $p_i(x,y) = x^i y^i$ при $i = 1, \ldots, n$. Эти многочлены линейно независимы. При этом

$$\langle p_i, p_j \rangle = \int_0^1 \int_0^1 x^{i+j} \cdot y^{i+j} \frac{dx}{x^{1-\alpha}} \cdot \frac{dy}{y^{1-\beta}} = \frac{1}{i+j+\alpha} \cdot \frac{1}{i+j+\beta},$$

а это в точности элементы исходной матрицы. Таким образом, это матрица скалярного произведения, и потому она положительно определена.

(*Л. Радзивиловский*)

О решении задачи 15.1

В «Математическом просвещении», вып. 23, опубликовано решение задачи 15.1. К сожалению, в пп. в) и г) и в упражнении 6 б) допущены погрешности, которые не влияют на ход рассуждений, но меняют ответ[2)]. Выражаем благодарность нашим читателям С. М. Лавренову и М. Б. Севрюку, заметившим эти погрешности. Ниже публикуем их письма.

В 23-м выпуске «Математического просвещения» на с. 226 приведена задача 15.1в:

Что больше: $\sqrt[3]{60}$ *или* $2+\sqrt[3]{7}$?

Задача имеет «простое» решение:

$$\sqrt[3]{60} > \frac{90}{23}, \qquad \text{так как } 60\cdot 23^3 - 90^3 = 1020 > 0.$$

$$\frac{90}{23} > 2+\sqrt[3]{7}, \quad \text{так как } (90-2\cdot 23)^3 - 23^3\cdot 7 = 15 > 0.$$

Ответ: $\sqrt[3]{60} > 2+\sqrt[3]{7}$.

Ответ на с. 227 неверный!

На с. 227 предлагается решение:

Заметим, что $\sqrt[3]{60} = 4\sqrt[3]{1-1/16}$ и $2+\sqrt[3]{7} = 2+2\sqrt[3]{1-1/8}$.

Разлагая оба выражения в ряд, имеем, что первые два главных члена совпадают, а третий член уже оказывается существенно больше у второго выражения (перед этим членом стоит знак «плюс») и это различие не может быть скомпенсировано остальными членами.

Попробуем аккуратно реализовать это решение.

$$f(x) = \sqrt[3]{1+x}, \quad f'(x) = \frac{1}{3(1+x)^{2/3}}, \quad f''(x) = -\frac{2}{9(1+x)^{5/3}}.$$

$$f(x) = 1 + \frac{1}{3}x + \frac{1}{2!}f'(\theta(x)x)x^2 = 1 + \frac{1}{3}x + \frac{1}{2}\cdot\left(-\frac{2}{9(1+\theta(x)x)^{5/3}}\right)x^2,$$

$$0 < \theta(x) < 1.$$

Остаточный член представлен в форме Лагранжа. Далее,

$$\sqrt[3]{60} = 4\sqrt[3]{1-\frac{1}{16}} = 4\left(1 - \frac{1}{3}\cdot\frac{1}{16} + \frac{1}{2}\left(-\frac{2}{9(1-\theta_1/16)^{5/3}}\right)\frac{1}{16^2}\right) =$$

$$= 4 - \frac{1}{12} - \frac{1}{9}\cdot\frac{1}{64}\cdot\frac{1}{(1-\theta_1/16)^{5/3}}, \quad \text{где } \theta_1 = \theta\left(\frac{1}{16}\right),$$

[2)] См. список опечаток в конце выпуска.

$$2+\sqrt[3]{7}=2+2\sqrt[3]{1-\frac{1}{8}}=2+2\Big(1-\frac{1}{3}\cdot\frac{1}{8}+\frac{1}{2}\Big(-\frac{2}{9(1-\theta_2/8)^{5/3}}\Big)\frac{1}{8^2}\Big)=$$
$$=4-\frac{1}{12}-\frac{1}{9}\cdot\frac{1}{32}\cdot\frac{1}{(1-\theta_2/8)^{5/3}},\quad\text{где }\theta_2=\theta\Big(\frac{1}{8}\Big).$$

Итак, первые два члена ряда действительно совпадают, но перед третьим членом стоит знак «минус», а не «плюс»!

Докажем, что

$$-\frac{1}{9}\cdot\frac{1}{64}\cdot\frac{1}{(1-\theta_1/16)^{5/3}}>-\frac{1}{9}\cdot\frac{1}{32}\cdot\frac{1}{(1-\theta_2/8)^{5/3}}.$$

После преобразований получаем

$$\Big(1-\frac{\theta_2}{8}\Big)^{5/3}<2\Big(1-\frac{\theta_1}{16}\Big)^{5/3}.$$

Возведём обе части в степень $3/5$ и получим после преобразований

$$\sqrt[5]{8}\cdot\theta_2-\sqrt[5]{32}\cdot\theta_1<16(\sqrt[5]{8}-1).$$

Заметим, что

$$\sqrt[5]{8}\cdot\theta_2-\sqrt[5]{32}\cdot\theta_1<\sqrt[5]{8}\cdot1-\sqrt[5]{32}\cdot0=\sqrt[5]{8}.$$

Но $\sqrt[5]{8}<16(\sqrt[5]{8}-1)$ (так как $15\sqrt[5]{8}>16\iff 6\,075\,000>1\,048\,576$). Неравенство доказано.

Впрочем, неравенство $15\sqrt[5]{8}>16$ можно доказать быстрее:

$$15\sqrt[5]{8}>16\quad\Leftrightarrow\quad\sqrt[5]{8}>\frac{16}{15}\quad\Leftrightarrow\quad 8>\Big(1+\frac{1}{15}\Big)^5.$$

Заметим, что $8>e>(1+1/15)^{15}>(1+1/15)^5$, что и требовалось.

В упражнении 6(б) ошибка в знаке неравенства. На самом деле

$$\sqrt[3]{413}=7{,}447\ldots>7{,}442\ldots=6+\sqrt[3]{3}.$$

Сергей Михайлович Лавренов,
ИПМ им. М. В. Келдыша, lasemi@mail.ru

В последнем выпуске 23 сборника «Математическое просвещение» (третья серия, 2019) допущена неаккуратность в решении задачи В. И. Арнольда 15.1.г) на с. 227. Приведённый в сборнике ответ

$$\int_0^{2\pi}\sin^{100}x\,dx\approx\frac{\sqrt{2}}{2}\quad\text{с погрешностью }\leqslant 20\,\%$$

неверен, правильный ответ

$$\int\limits_0^{2\pi} \sin^{100} x\,dx \approx \frac{1}{2}$$

(кстати, сам В. И. Арнольд в статье «Математический тривиум» // УМН, 1991, т. 46, вып. 1, с. 225–232 на с. 226 говорил о точности в 10 %).

Действительно, как справедливо пишет автор решения А. Я. Белов,

$$\int\limits_0^{2\pi} \sin^{100} x\,dx = \int\limits_0^{2\pi} \cos^{100} x\,dx \approx 2\int\limits_{-\infty}^{+\infty} e^{-50x^2}\,dx = 2\sqrt{\frac{\pi}{50}}. \qquad (1)$$

Заменяя $\pi = 3{,}14159\ldots$ приближённым значением $25/8 = 3{,}125$, получаем $1/2$. (При публикации здесь была допущена арифметическая ошибка.)

В этой задаче есть дополнительный нюанс: все интегралы

$$A_n = \int\limits_0^{2\pi} |\sin^n x|\,dx = 4\int\limits_0^{\pi/2} \sin^n x\,dx, \quad n = 0, 1, 2, \ldots,$$

можно очень просто вычислить *точно*, причём без явного использования напрашивающегося здесь аппарата бета- и гамма-функций. Действительно, при $n \geqslant 2$ имеем

$$A_n = -4\int\limits_0^{\pi/2} \sin^{n-1} x\,d(\cos x) = 4\int\limits_0^{\pi/2} \cos x\,d(\sin^{n-1} x) =$$

$$= 4(n-1)\int\limits_0^{\pi/2} \sin^{n-2} x\,(1-\sin^2 x)\,dx = (n-1)(A_{n-2} - A_n),$$

так что

$$A_n = \frac{n-1}{n} A_{n-2}, \quad n \geqslant 2. \qquad (2)$$

Так как $A_0 = 2\pi$ и $A_1 = 4$, то при $k \geqslant 0$

$$A_{2k} = 2\pi\frac{(2k-1)!!}{(2k)!!}, \quad A_{2k+1} = 4\frac{(2k)!!}{(2k+1)!!}.$$

Подчеркнём, что это *точные* равенства. Теперь

$$A_{100} = 2\pi\frac{99!!}{100!!} = 2\pi\frac{100!}{(100!!)^2} = 2\pi\frac{100!}{(2^{50}\,50!)^2},$$

и по формуле Стирлинга

$$A_{100} \approx 2\pi\sqrt{200\pi}\left(\frac{100}{e}\right)^{100}\frac{1}{100\pi\cdot 2^{100}}\left(\frac{e}{50}\right)^{100} = \frac{\sqrt{2\pi}}{5},$$

что, как и следовало ожидать, совпадает с прежним значением (1).

Арнольд почти наверняка имел в виду решение Белова (идея замены $\cos x$ на $e^{-x^2/2}$ в окрестности нуля применима к целому классу интегралов), а не изложенное выше альтернативное решение, опирающееся на соотношение (2).

Прямое компьютерное вычисление даёт

$$A_{100} = 2\pi \prod_{k=1}^{50} \frac{2k-1}{2k} = 0{,}5000739\ldots,$$

так что $1/2$ приближает A_{100} с погрешностью всего 0,015 %! Интересно, что $25/8$ приближает π с погрешностью 0,53 %.

Михаил Борисович Севрюк,
Институт энергетических проблем
химической физики РАН им. В. Л. Тальрозе,
2421584@mail.ru, sevryuk@mccme.ru

Опечатки, замеченные в выпуске 13

Страница,	строка	Напечатано	Следует читать
180	10 снизу	$\leqslant$	$\geqslant$

Опечатки, замеченные в выпуске 16

Страница,	строка	Напечатано	Следует читать
230	1 снизу	неравенство Виртингера	Фольклор

Опечатки, замеченные в выпуске 23

Страница,	строка	Напечатано	Следует читать
213	4 снизу	arsh	arch
227	14 сверху	$<$	$>$
227	17 сверху	больше	меньше
227	18 сверху	плюс	минус
228	6 снизу	больше	меньше

Информация для авторов

1. Сборник «Математическое просвещение» предназначен для широкого круга научных работников, преподавателей, учащихся и всех, кто интересуется математикой. Издание публикует материалы по различным областям математики, а также по проблемам её истории и преподавания, интересные и доступные указанной аудитории.

2. Сборник «Математическое просвещение» не публикует существенно новые научные результаты, оценка которых доступна лишь специалистам в соответствующей области. Не публикуются также материалы по текущим вопросам преподавания математики в учебных заведениях.

3. Материалы принимаются по электронной почте на адрес matpros@yandex.ru в виде двух файлов (pdf и tex) с дополнительными файлами рисунков и т. п., если требуется. Допускается присылка статей, набранных в Word.

4. Просим обратить внимание, что материалы принимаются в чёрно-белом исполнении.

5. Просим авторов кратко пояснять в начале статьи, в чём её цель и почему тема статьи представляет интерес.

6. Редакция благодарна авторам за оформление ссылок на литературу как в предыдущих выпусках, см. http://www.mccme.ru/free-books/matpros.html

7. В конце статьи необходимо указать для каждого из авторов:

— фамилию, имя, а также отчество (если есть) полностью,

— место работы/обучения,

— электронный адрес для публикации.

8. Авторы задач вместе с условием представляют письменное решение (хотя бы набросок).

9. Авторы опубликованных статей имеют право на 2 экземпляра сборника каждый, просим обращаться по адресу matpros@yandex.ru

Научно-популярное издание

Издательство Московского центра
непрерывного математического образования

119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241–08–04.

Отпечатано в ООО «Типография „Миттель-Пресс“».
г. Москва, ул. Руставели, д. 14, стр. 6.
Тел./факс +7 (495) 619–08–30, 647–01–89.
E-mail: mittelpress@mail.ru

Подписано в печать 15.03.2019 г. Формат $70\times100\,{}^{1}/_{16}$. Бумага офсетная.
Печать офсетная. Печ. л. 13. Тираж 800 экз. Заказ №