МАТЕМАТИЧЕСКОЕ ПРОСВЕЩЕНИЕ

Третья серия

выпуск 20

Москва
Издательство МЦНМО
2016

УДК 51.009
ББК 22.1
М34

Редакционная коллегия

Бугаенко О. В.	Дориченко С. А.	Райгородский А. М.
Винберг Э. Б.	Егоров А. А.	Розов Н. Х.
Вялый М. Н.	Заславский А. А.	Сосинский А. Б.
Гайфуллин А. А.	Ильяшенко Ю. С.	Тихомиров В. М.
Гальперин Г. А.	Канель-Белов А. Я.	Устинов А. В.
Глейзер Г. Д.	Константинов Н. Н.	Френкин Б. Р.
Гусейн-Заде С. М.	Прасолов В. В.	Ященко И. В.

Главный редактор Э. Б. Винберг
Отв. секретарь Б. Р. Френкин

Адрес редакции:

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

М34 **Математическое просвещение.** Третья серия, вып. 20. —
М.: МЦНМО, 2016. — 272 с.
ISBN 978-5-4439-1029-1

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

ISBN 978-5-4439-1029-1

© МЦНМО, 2016.

Содержание

Математический мир

С. С. Демидов, С. С. Петрова, В. М. Тихомиров, Т. А. Токарева
К 150-летию Московского математического общества 5

В. М. Тихомиров
Александр Михайлович Абрамов (5 июня 1946 г. – 24 мая 2015 г.) . . . 18

В. А. Гурвич
Торг с богами 23

Геометрия: классика и современность

Р. Р. Пименов
Тройственная симметрия фрактального калейдоскопа 57

Г. Гальперин, Ч. Делман
Повесть о трёх кругах 111

Наш семинар: математические сюжеты

А. Б. Певный, С. М. Ситник
Средние Джини 135

Н. В. Филимоненкова, П. А. Бакусов
Гиперболические многочлены и конусы Гординга 143

В. М. Журавлёв, П. И. Самовол
Экспоненциальные диофантовы уравнения и сумма цифр числа 167

Ф. А. Шаров
Разрезания прямоугольника на прямоугольники с заданными отношениями сторон 200

По мотивам задачника

Л. Радзивиловский, Д. Радченко, М. Танцюра, И. Фещенко
Разрезание параллелепипеда на бруски 215

А. М. Райгородский
Графы с большим хроматическим числом и большим обхватом 228

М. Д. Турбин
Задача Майхилла о синхронизации ряда стрелков 238

Нам пишут

А. Ю. Юрьев
Случайные блуждания возвращаются 243

А. Д. Матушкин
Непустота пересечения цепочки множеств 247

Задачник

Условия задач 249

Решения задач из прошлых выпусков 252

Указатель условий, решений и статей по мотивам задач из «Математического просвещения» 269

Математический мир

К 150-летию Московского математического общества

С. С. Демидов, С. С. Петрова, В. М. Тихомиров, Т. А. Токарева

В 2014–2017 гг. российское математическое сообщество отмечает 150-летие возникновения своей первой общественной организации — Московского математического общества — и её печатного органа: журнала «Математический сборник». Появившись на свет в 1864 году, Общество уже в первые годы своего существования проявило себя как явление общенационального характера. Как писал А. П. Юшкевич в своём фундаментальном труде о развитии математики в России [20, стр. 317], «по своему значению Московское математическое общество уступало только Академии наук». «Событие, которое мы отмечаем, — читаем мы в докладе ректора Московского университета [14] на торжественном заседании, посвящённом этому юбилею, — по своему значению гораздо шире, чем просто юбилей Московского математического общества — это полтора века российской науки, российской математики, это полтора века Московского университета».

1. Рождение Московского математического общества. Первая попытка организации математического общества в Москве была предпринята группой преподавателей и студентов университета ещё в 1810 г. [17]. Однако просуществовало оно недолго: тогдашняя первопрестольная ещё не обладала достаточным количеством активных профессиональных математиков, способных поддерживать его нормальную деятельность. Необходимые для этого условия сложились лишь к 1860-м годам, когда в Москве сформировался математический центр европейского уровня. Его созданию способствовала сама атмосфера, царившая в российском обществе того

времени — в период судьбоносных преобразований в социальной жизни, случившихся в России после восшествия на трон в 1855 г. Александра II. Самой известной из тогдашних реформ стала отмена крепостного права.

Из преобразований, имевших прямое отношение к рассматриваемому нами вопросу, назовём фундаментальные изменения в организации системы народного образования, в частности, образования гимназического и высшего. Принятый в 1863 г. новый университетский устав увеличил число позиций для представителей математических наук, а также содержал рекомендации об учреждении при университетах научных обществ. В 2014 г. можно было отпраздновать 150-летие принятия Гимназического устава. Об уровне реформ, последовавших за отменой крепостного права, свидетельствует следующий фрагмент Гимназического устава: «Гимназия имеет целью, посредством правильного воспитания и обучения общеобразовательным наукам, развить молодых людей в умственном и нравственном отношениях так, чтобы они могли или с успехом начать своё специальное образование в одном из высших учебных заведений, или прямо приступить к полезной деятельности на избранном ими поприще общественной жизни».

В этой атмосфере в 1864 г. при Московском университете было создано математическое общество. Первое его заседание было собрано 15 (27) сентября на квартире инициатора создания общества профессора Н. Д. Брашмана, состоявшего в то время на пенсии и по состоянию здоровья уже не имевшего возможности посещать университет. Он же был избран его президентом. Вице-президентом Общества стал А. Ю. Давидов, а секретарём В. Я. Цингер.

Поначалу Общество преследовало в своей деятельности очень скромные цели. Как следует из протокола его первого заседания, «цель Общества есть взаимное содействие в занятиях математическими науками» [12, стр. 472]. Однако вскоре выяснилось, что лидеры Общества ставили перед ним более амбициозные цели. Так, представляя в 1866 г. документы (и среди них проект нового устава) в Министерство народного просвещения, они так сформулировали стоявшие перед Обществом задачи: «Московское математическое общество учреждается с целью содействовать развитию математических наук в России» [18, стр. III].

На четвёртом заседании Общества 15 декабря 1864 г. его руководители выразили мнение, что доклады, прочитанные на заседаниях, заслуживают публикации. И уже в апреле 1865 г. было принято решение об издании журнала. При обсуждении вопроса о языке, на котором будет выходить журнал, победила точка зрения Н. В. Бугаева, считавшего, что для развития математики в стране целесообразно печатать его по-русски. Первый том журнала, названного «Математическим Сборником», появился в октябре 1866 г. и был посвящён памяти скончавшегося в мае того же

года первого президента Общества Н. Д. Брашмана. Так началось издание одного из самых влиятельных математических журналов XX века [3].

Государем императором Московское математическое общество было утверждено 28 января 1867 г. С этого момента Общество начало своё существование не только де-факто, но и де-юре. Тогда же был утверждён и новый Устав, в первом параграфе которого были указаны его цели (см. выше), а согласно второму определено, что членами «Общества могут быть доктора и магистры русских университетов по математическим наукам и вообще лица, известные Обществу своими учеными трудами в области этих наук» [18, стр. III]. Примечанием же к § 5 свидетельствовалось, что при открытии «Общества его действительными членами становятся гг. учредители: А. Ю. Давидов, Ф. А. Бредихин, Ф. А. Слудский, В. Я. Цингер, М. Ф. Хандриков, Н. В. Бугаев, Н. Н. Алексеев, А. В. Летников, кн. С. С. Урусов, К. М. Петерсон, Е. Ф. Сабинин и С. А. Юрьев» [18, стр. IV]. Среди них мы видим преподавателей как Московского университета (астроном Бредихин, математик Бугаев, механик Слудский, физик Н. А. Любимов), так и других учебных заведений: профессор Московского технического училища А. В. Летников, скромный преподаватель математики немецкой гимназии (Московского Петропавловского училища) К. М. Петерсон. 18 февраля 1867 г. на первом после официального утверждения Общества заседании было принято решение добавить к числу его членов-учредителей, количество которых стало таким образом 14, Брашмана, первого президента Общества, и проживавшего в Петербурге академика П. Л. Чебышёва, роль которого при организации и начальном этапе существования Общества была очень существенной.

Так началась история Московского математического общества, одного из старейших в мире. Лондонское математическое общество, например, было основано в 1865 г., Математическое общество Франции «было основано по образу Лондонского и Московского» [16, стр. XII] в 1872 г., Математический кружок в Палермо — в 1884 г., Нью-Йоркское математическое общество (впоследствии преобразованное в Американское математическое общество) — в 1888 г., Германское объединение математиков — в 1890 г. Старше всех, насколько нам известно, Голландское математическое общество, основанное в 1778 г.

2. Юность Общества. Последняя треть XIX – начало XX века — период активного развития Общества. И если в числе его учредителей насчитывалось всего 14 человек, то в начале XX века оно объединяло уже 101 математика, а к началу Первой мировой войны его состав достиг 112 человек, из которых только 34 были москвичами, а 57 представляли другие города России, 21 — зарубежные ученые. Таким образом, деятельность Общества принимала не только общероссийский, но и международный характер.

К началу XX века Москва выросла в один из заметных в научной жизни Европы центров математических исследований, известный, прежде всего, своими научными школами по прикладной математике (Н. Е. Жуковский, С. А. Чаплыгин) и дифференциальной геометрии (Б. К. Млодзеевский, Д. Ф. Егоров), а также результатами в области проективной геометрии (К. А. Андреев, А. К. Власов), теории чисел (Н. В. Бугаев), теории функций комплексного переменного (П. А. Некрасов).

Становление московской математической школы проходило в условиях устойчивой конфронтации с другой, значительно более известной тогда в Европе — петербургской школой П. Л. Чебышёва (А. А. Марков, А. М. Ляпунов и др.). Основанием для этой конфронтации стали, прежде всего, разногласия идейного толка, определившие до известной степени математическую ориентацию обеих школ: позитивизм, либеральный демократизм, воинствующий атеизм и антимонархизм, доминировавшие в стане петербургских математиков, с одной стороны; антипозитивизм, увлеченность идеалистической и даже религиозной философией, православные настроения и монархизм москвичей, — с другой.

Это противостояние накладывало отпечаток на жизнь всего российского математического сообщества последней трети XIX – первой трети XX века, создав в нем известное напряжение. (Не надо забывать, что подавляющую часть профессуры провинциальных университетов — Харьковского, Киевского, Новороссийского, Варшавского, Юрьевского — составляли воспитанники Московского и Петербургского университетов.) Это напряжение приводило к конфликтным ситуациям, зачастую заканчивавшимся открытыми столкновениями, как это случилось, например, в баталиях по поводу работ 1887–1891 гг. академика В. Г. Имшенецкого, прибегшего к поддержке москвичей в своем конфликте с А. А. Марковым и А. Н. Коркиным; или в связи со знаменитыми результатами 1888 г. С. В. Ковалевской, касавшимися задачи движения тела с одной неподвижной точкой, на которые в 1890-е годы ополчился все тот же А. А. Марков; наконец, в ожесточенной полемике последнего с П. А. Некрасовым по вопросам теории вероятностей и ее преподавания в средней школе. Все эти конфликты обсуждались на заседаниях Московского математического общества, которое вынуждено было выступать в роли арбитра.

Здесь со всей отчетливостью проявилась особая роль Общества, столь важная в период становления российского математического сообщества — роль его как *общественной* организации. Притом организации, деятельность которой носила не местный (московский!), а общенациональный характер.

Разумеется, особо важную роль играло Общество в развитии математических исследований в самой Москве, начавшей вырастать из просто

заметного в Европе центра математических исследований в одну из ведущих мировых математических столиц.

Московское математическое общество стало той площадкой, на которой зарождалась московская математическая школа.

Устное общение московских математиков вплоть до начала десятых годов XX века проходило в основном на заседаниях Общества. Затем наступила эра семинаров. Начало было положено Д. Ф. Егоровым, а затем особую роль в рождении московской математической школы сыграли личность Н. Н. Лузина и его семинар, столетие которого тоже отмечалось в 2014 г. А на заседаниях Общества помимо основоположников новых направлений — Егорова и Лузина — стали выступать и соратники Лузина — В. В. Голубев, И. И. Привалов и В. В. Степанов, и первое поколение лузинских учеников: П. С. Александров, Д. Е. Меньшов, М. Я. Суслин, Ф. Я. Хинчин и П. С. Урысон.

3. Период испытаний. Первая мировая война, последовавшие за ней революционные события и гражданская война расстроили математическую жизнь России. Решающая роль в ее возобновлении принадлежала возглавившему с 1923 г. Московское математическое общество директору Научно-исследовательского института математики и механики при I МГУ (НИИММ МГУ) Д. Ф. Егорову. До него президентами Общества были: Н. Д. Брашман (1864–1866), А. Ю. Давидов (1866–1886), В. Я. Цингер (1886–1891), Н. В. Бугаев (1891–1903), П. А. Некрасов (1903–1905), Н. Е. Жуковский (1905–1921), Б. К. Млодзеевский (1921–1923)).

В 1924 г. под редакцией Егорова после пятилетнего перерыва вышел 31-й том «Математического сборника» (первые выпуски 30-го тома начали издаваться в 1916 г., а последний увидел свет лишь в 1918 г.). Желая покончить со сложившейся конфронтацией двух ведущих математических школ, москвичи ввели в редколлегию 32-го тома В. А. Стеклова. В возобновленном журнале стали охотнее, чем раньше, печататься петроградцы-ленинградцы: в 31-м томе — А. С. и Я. С. Безиковичи, Г. М. Фихтенгольц и В. А. Фок, в 31-м и 42-м — И. М. Виноградов, в 32-м и 35-м — Н. М. Гюнтер, в 34-м — И. А. Лаппо-Данилевский, в 38-м и 40-м — С. Л. Соболев, а в 41-м томе Л. В. Канторович. В конце 1920-х – 1930-х гг. журнал занял позиции центрального всесоюзного журнала. На его страницах помещали свои статьи математики из разных научных центров СССР — Казани (Н. Г. Чеботарев), Киева (Д. А. Граве, Н. М. Крылов), Ташкента (В. И. Романовский), Ростова-на-Дону (Д. Д. Мордухай-Болтовской), Одессы (М. Г. Крейн) [3].

С целью восстановления международных научных связей была изменена редакционная политика «Сборника» — в нем стали публиковаться работы иностранных математиков, снабженные резюме на русском, в то

время как статьи отечественных ученых сопровождались в нем аннотациями на наиболее распространенных европейских языках. Эта начинание оказалось успешным. Авторами «Математического сборника» второй половины 1920-х годов стали: Э. Картан (Т. 34, 42), М. Фреше (Т. 32), Ж. Адамар (Т. 41), Х. Хопф (Т. 37), С. Лефшец (Т. 39), Р. Мизес (Т. 41), Э. Нётер (Т. 36), Р. Мемке (Т. 36), В. Серпинский (Т. 31, 36), Л. Тоннелли (Т. 33) [3].

По инициативе Общества, НИИММ МГУ и при активном организационном участии Д. Ф. Егорова весной 1927 г. был созван первый Всероссийский математический съезд, заложивший организационные основы всей последующей математической жизни в СССР [16].

В результате инициатив и усилий Д. Ф. Егорова Московское математическое общество ко второй половине 1920-х годов заняло ведущие позиции в отечественной математике.

4. Советская власть и математическое сообщество. Одной из важнейших задач советской власти в 1920-е годы стало наведение «порядка» в науке и образовании. В противовес Российской академии наук в 1918 г. создается Коммунистическая (до 1924 г. — Социалистическая) академия, призванная содействовать развитию естественных и общественных наук на основе марксистской методологии. Сама Российская академия наук в 1925 г. переименовывается в Академию наук СССР и становится «высшим ученым учреждением СССР». Новый Устав академии (1927), утвержденный Политбюро ЦК ВКП(б), ставит перед АН СССР задачу — «приспосабливать научные теории... к практическому применению в промышленности и культурно-экономическом строительстве Союза ССР».

Приверженность «чистой» науке, отрицательное отношение к марксистской идеологии, религиозность сделали избранного в 1929 г. почетным членом АН СССР лидера московских математиков Д. Ф. Егорова объектом нападок со стороны группирующихся вокруг Комакадемии математиков. В конце 1929 г. он был смещен с поста директора НИИММ МГУ, а в октябре 1930 г. арестован по делу о так называемой истинно-православной церкви и сослан в Казань, где скончался в 1931 г. [19].

Под ударом оказалось Московское математическое общество, и, чтобы его «спасти», группа молодых его членов (Л. А. Люстерник, Л. Г. Шнирельман, А. О. Гельфонд, Л. С. Понтрягин и К. П. Некрасов) выступила с «Декларацией», в которой говорилось, что «Общество должно быть коренным образом реорганизовано. Оно должно очистить свои ряды от контрреволюционеров, оно должно разорвать кастовые перегородки, стать массовым, слиться с пролетарской общественностью, стать организатором науки в условиях социалистического строительства» [11, стр. 70]. В результате президентом Общества был «избран» работавший в Отделе агитации

и пропаганды ЦК ВКП(б) видный партийный функционер и совсем слабый математик — Э. Кольман. Правда, президентом он пробыл недолго (в 1930–1932 гг.). Да и непосредственные свидетели этих событий считали, что с конца 1930 и весь 1931 год Общество «фактически бездействовало» [9, стр. 21], [2, стр. 11]. Уже в 1932 г. на пост президента был избран П. С. Александров, сохранявший его вплоть до 1964 г.

«Дело» Егорова ознаменовало новый этап в жизни отечественного математического сообщества — отныне в его среде и тем более в его руководстве не терпелось никакое инакомыслие по вопросам политики или идеологии.

5. Рождение советской математической школы. В конце двадцатых — начале тридцатых годов двадцатого столетия многие выдающиеся московские математики (П. С. Александров, А. Н. Колмогоров, Д. Е. Меньшов, М. А. Лаврентьев, Л. А. Люстерник, А. Я. Хинчин, Л. Г. Шнирельман, О. Ю. Шмидт) побывали в командировках за границей (в Париже, Гёттингене и других местах), где они докладывали о своих выдающихся результатах. Наступил первый золотой период развития советской математики.

На XVII съезде ВКП(б) (1934 г.) в докладе народного комиссара просвещения РСФСР А. С. Бубнова были отмечены два выдающихся результата математиков А. О. Гельфонда и Л. Г. Шнирельмана. В июне 1934 г. в Ленинграде состоялся II Всесоюзный математический съезд, констатировавший «взаимное сближение» основных математических школ Советского Союза и, главное, двух ведущих — Ленинградской-Петербургской и Московской, что знаменовало рождение единой советской математической школы [6, 7, 16].

В тридцатые годы при активном участии Московского математического общества состоялись две выдающиеся международные конференции, организованные В. Ф. Каганом и П. С. Александровым, — по дифференциальной геометрии и топологии.

Перевод Академии наук СССР в Москву (1934 г.) предполагал переезд в столицу большинства ее институтов, в том числе Физико-математического института, на базе которого был образован Математический институт им. В. А. Стеклова АН СССР (МИАН). Это окончательно объединило петербургских-ленинградских и московских математиков. Сложился уникальный по широте деятельности и концентрации математических талантов центр. Произошел синтез традиций петербургской школы математической физики (С. Л. Соболев) и московской, восходящей к К. М. Петерсону, геометрической традиции исследований в области теории дифференциальных уравнений с частными производными (И. Г. Петровский); чебышевской линии развития теории вероятностей (С. Н. Бернштейн) с московской, выросшей в недрах метрической теории функций (А. Я. Хинчин, А. Н. Колмогоров); встретились две линии развития алгебраических исследований,

восходящих к киевской школе Д. А. Граве, — ленинградская (Б. Н. Делоне) и московская (О. Ю. Шмидт, А. Г. Курош).

В результате МИАН, механико-математический факультет МГУ и Московское математическое общество составили тот научный потенциал, который способствовал превращению советской математической школы в одну из ведущих математических школ.

Вторая половина тридцатых годов была омрачена многими трагическими событиями, происходившими в нашей стране. Это отразилось и на развитии математической науки.

Железный занавес прервал продуктивные научные контакты советских математиков с зарубежными коллегами.

6. Московское математическое общество после Великой Отечественной войны. Великая Отечественная война и вынужденная эвакуация осенью 1941 г. МИАН в Казань, а МГУ в Среднюю Азию и затем в Свердловск не прервали работу Московского математического общества. Возникло два его отделения: Казанское, которым руководил П. С. Александров, и Ташкентское, работавшее под руководством вице-президента Общества В. В. Степанова. Осенью 1943 г., после реэвакуации института и университета, отделения Общества объединились [15].

В послевоенный период Московское математическое общество продолжает занимать совершенно особое место в математическом сообществе. Находясь в столице, оно остается в центре главных событий — на его заседаниях докладывались практически все наиболее значимые результаты, полученные отечественными математиками. Прерванные войной научные коммуникации восстанавливаются, привлекая ведущих зарубежных математиков к участию в созываемых Обществом съездах, конференциях, семинарах. Расширяется и издательская деятельность Общества. Кроме периодических изданий — «Математического сборника», «Успехов математических наук» (с 1946 г.) и почти периодических — «Трудов Московского математического общества» (с 1952 г.), «Трудов семинара имени И. Г. Петровского» (с 1976 г.) — им публикуются такие фундаментальные работы, как «Математика в СССР за 30 (за 40) лет», «Математика в СССР. 1958–1967. Т. II.» и др. С 1945 г. возобновляется присуждение премий Общества. (Премии были учреждены в середине 1930-х годов для молодых ученых до 30 лет, работы которых были доложены на заседаниях Московского математического общества.)

В пятидесятые годы наступил второй золотой период в жизни Московского математического общества, Московского университета и механико-математического факультета. В подтверждение этого приведём слова В. И. Арнольда.

«Плеяда великих математиков, собранных на одном факультете, представляла собой явление совершенно исключительное, и мне не приходилось встречать ничего подобного более нигде.

Колмогоров, Гельфанд, Петровский, П. С. Новиков, Марков, Гельфонд, Люстерник, Хинчин и П. С. Александров учили таких студентов, как Манин, Синай, С. Новиков, Алексеев, Аносов, Кириллов и я сам».

О необычайном расцвете математики в период пятидесятых-шестидесятых годов свидетельствует уровень докладов на Математическом обществе в конце 1969 – начале 1970 года (в скобках указано число посетителей доклада).

18.11.69. И. М. Гельфанд, Д. Б. Фукс. Когомологии бесконечных алгебр Ли (158).

25.11.69. С. И. Адян. Тождественные соотношения в группах (72).

02.11.69. Г. Н. Тюрина. Универсальные деформации изолированных особенностей комплексных пространств (82).

25.01.70. Э. Колчин. Теория Галуа дифференциальных уравнений (94).

17.02.70. В. А. Белинский, Е. М. Лифшиц, И. М. Халатников. Об особенностях когомологических решений уравнений гравитации (91).

24.02.70. А. Н. Колмогоров. Статистика гидродинамики океана (139).

03.03.70. В. А. Рохлин. Двумерные гомологии и двумерные подмногообразия четырёхмерных многообразий (122).

В период с 1955 г. по 1965 г. премии Московского математического общества получили Р. Л. Добрушин и Ф. И. Карпелевич (1955), С. И. Адян и А. Г. Витушкин (1956), В. И. Арнольд и Ф. А. Березин (1957), А. А. Мучник и А. С. Шварц (1958), В. А. Кондратьев и О. Б. Лупанов (1959), А. Г. Костюченко — Б. С. Митягин, Я. Г. Синай (1960), О. В. Бесов (1961), А. С. Дынин, А. Л. Онищик, Л. А. Сахнович, Е. Г. Скляренко (1962), А. Д. Вентцель, Э. Б. Винберг — С. Г. Гиндикин, Ю. И. Манин (1963), Е. С. Голод, С. П. Новиков, Е. П. Долженко (1964), Д. В. Аносов, А. А. Кириллов (1965).

В 1964 г. в Московском университете широко отмечалось столетие старейшего математического общества страны. Этой дате было посвящено торжественное заседание Общества, состоявшееся 20 октября в актовом зале МГУ. Накануне, 19 октября, статью «Московское математическое общество» (написанную А. Н. Колмогоровым и А. Г. Курошем) опубликовала центральная газета страны — «Правда» [8], открыв тем самым список приветствий по случаю юбилея, к которому присоединились многие научные учреждения, университеты, советские и зарубежные общества и отдельные ученые.

Со вступительным докладом на юбилейном торжестве выступил Павел Сергеевич Александров [1], возглавлявший Московское математическое общество в течение 32 лет. На этом же заседании о деятельности Общества

за последнее тридцатилетие его вековой истории рассказал Александр Геннадиевич Курош [9]. Своими воспоминаниями о жизни Общества в 10-е и 20-е годы XX века поделились Дмитрий Евгеньевич Меньшов [13] и Лазарь Аронович Люстерник [10].

В прошлом году Общество отметило свой 150-летний юбилей конференцией «Московское математическое общество и Московский государственный университет имени М. В. Ломоносова», которая прошла 23–25 декабря 2014 года на механико-математическом факультете Московского университета. Конференцию открыл действующий президент Общества В. А. Васильев. Затем был прочитан доклад ректора Московского университета В. А. Садовничего [20]. С докладом «Сингулярные солитоны» выступил почётный президент Общества С. П. Новиков. Вице-президент РАН В. В. Козлов сделал доклад «Динамические системы и статистическая механика». С приветствием от Санкт-Петербургского математического общества выступил А. М. Вершик. Наконец об истории Общества рассказали С. С. Демидов и В. М. Тихомиров (расширенный текст см. [5]). Конференция продолжилась 24–25 декабря докладами по актуальным проблемам математики и вопросам её преподавания.

Значимость Московского математического Общества в жизни отечественной математики подчёркивается тем фактом, что его президентами избирались крупнейшие учёные. После Александрова президентами Общества становились А. Н. Колмогоров (1964–1966), И. М. Гельфанд (1966–1970), И. Р. Шафаревич (1970–1973), А. Н. Колмогоров (1973–1985), С. П. Новиков (1985–1996), В. И. Арнольд (1996–2010), В. А. Васильев (2010 — н. вр.).

И сегодня Общество по-прежнему собирается по вторникам на свои заседания, где заслушиваются сообщения о новых результатах и обзорные доклады, обсуждаются проблемы, стоящие перед математическим сообществом. Общество по-прежнему волнуют проблемы общенационального характера.

Одной из них остается постановка преподавания математики, прежде всего в средней школе. Эта тема находится в центре внимания Общества с момента его основания. В первых томах «Математического сборника» существовал даже специальный раздел, предназначенный для учителей гимназий. В феврале 1934 г. была создана секция элементарной математики, вскоре переименованная в научно-педагогическую, а затем в секцию средней школы. Во время войны секция не функционировала, возобновив свою работу лишь в 1948 г. (под руководством А. И. Маркушевича). Особенно следует отметить инициативу Общества середины 1930-х годов, связанную с организацией школьных математических олимпиад, первая из которых была проведена осенью 1935 г.

На протяжении своей истории Общество неоднократно обсуждало проблемы, связанные с программами и учебниками по математике для средней школы. Так, 27 ноября 2001 г. прошло заседание Общества, посвящённое планируемой Министерством образования РФ реформе школы и перспективам математического образования в стране. Планы министерства вызвали негативную реакцию в Обществе. Заседание было многолюдным и чрезвычайно эмоциональным.

7. Заключение. Московское математическое общество возникло в пореформенной России в период экономического и культурного подъема, сопровождавшего эпоху правления Александра II. На протяжении всей своей полуторавековой славной истории оно, будучи достаточно консервативным в вопросах организации и форм деятельности, тем не менее, очень чутко реагировало и отвечало на запросы времени.

Основанное как кружок математиков, поддерживающих друг друга в научных занятиях, оно трансформировалось в неформальный институт, в значительной степени координирующий и организующий деятельность национального математического сообщества. Общество с переменным успехом играло эту роль вплоть до 30-х годов XX века. Перестав быть таковым в связи с советской реорганизацией национального математического сообщества, сопровождавшейся передачей управления им Академии наук СССР, а также Министерству высшего и среднего специального образования, Московское математическое общество взяло на себя роль ведущей в стране общественной, до известной степени независимой от партийных и государственных органов, организации математиков, которая успешно преодолевала встречавшиеся на ее пути рифы, сохраняя высокую научную репутацию и авторитет.

Возможно, именно эти достоинства помогли Московскому математическому обществу справиться и с перестроечными неустройствами. Этому же способствовал и его статус вольного научного общества. Ему не были нужны серьезные финансовые вливания, и оно в меньшей степени оказалось в зависимости от оттока за рубеж высококвалифицированных научных кадров. Даже уехав, члены Общества не порывают с ним связей, продолжая печатать в изданиях Общества свои работы и выступая на его заседаниях с докладами во время своих кратковременных визитов на родину. Как уже случалось в его долгой истории, на которую приходились и годы процветания, и годы разрухи, оно, несмотря ни на что, продолжает свою деятельность.

И это более всего вселяет оптимизм и веру в продолжение математических исследований в России, о чем всегда пеклось Московское математическое общество. На столетнем юбилее Общества Павел Сергеевич Александров произнес следующие слова:

«...Московское математическое общество всегда культивировало... многогранное развитие математики, не стараясь втиснуть его ни в какие заранее данные рамки и системы оценок. В течение десятилетий Московское математическое общество было тем местом, на котором произрастали и жили математические открытия, искания, волнения, все творческие эмоции московских математиков нескольких поколений. Московское математическое общество не было только местом, где регистрировались отдельные математические результаты, где читались популярные лекции по математике. Московское математическое общество было школой математической эстетики, математического вкуса, очень взыскательного, и школой математической этики, научной этики, тоже очень взыскательной...» [1, стр. 9].

Список литературы

[1] *Александров П. С.* Вступительный доклад на торжественном заседании Московского математического общества 20 октября 1964 г. // УМН. 1965. Т. XX, вып. 3(123). С. 4–9.

[2] *Александров П. С.* и *Головин О. Н.* Московское математическое общество // УМН. 1957. Т. XII, вып. 6(78). С. 9–46.

[3] *Демидов С. С.* «Математический сборник» в 1866–1935 гг. // Историко-математические исследования. Вторая серия. 1996. Вып. 1(36), № 2. С. 127–145.

[4] *Демидов С. С.* Профессор Московского университета Д. Ф. Егоров и имеславие в России // Историко-математические исследования. Вторая серия. 1999. Вып. 4(39). С. 123–156.

[5] *Демидов С. С., Тихомиров В. М., Токарева Т. А.* Московское математическое общество в развитии отечественной математики (к 150-летию основания) // УМН. 2015. Т. 70, вып. 1(421). С. 189–203.

[6] *Демидов С. С., Токарева Т. А.* Рождение Советской математической школы // Наука и техника в первые десятилетия советской власти: социокультурное измерение (1917–1940) / Отв. ред. Е. Б. Музрукова. М., 2007. С. 347–375.

[7] *Демидов С. С., Токарева Т. А.* Формирование Советской математической школы // Историко-математические исследования. Вторая серия. 2005. Вып. 10(45). С. 142–159.

[8] *Колмогоров А. Н., Курош А. Г.* Московское математическое общество. К столетней годовщине // Правда. 1964. 19 октября. С. 4.

[9] *Курош А. Г.* Московское математическое общество за последнюю треть века. (Доклад на юбилейном заседании Общества 20.10.1964.) // УМН. 1965. Т. XX, вып. 3(123). С. 10–18.

[10] *Люстерник Л. А.* Выступление на заседании Московского математического общества // УМН. 1965. Т. XX, вып. 3(123). С. 21–30.

[11] *Люстерник Л. А., Шнирельман Л. Г., Гельфонд А.* [*О.*], *Понтрягин Л.* [*С.*] *Некрасов.* Декларация инициативной группы по реорганизации Математического общества // Научный работник. 1930. № 11–12. С. 67–71.

[12] Материалы для истории Московского математического общества // Матем. сб. 1889. Т. XIV, вып. 3. С. 471–486.

[13] *Меньшов Д. Е.* Московское математическое общество в период 1910–1920 гг. // УМН. 1965. Т. XX, вып. 3(123). С. 19–20.

[14] *Садовничий В. А.* Московское математическое общество и Московский университет. http://letopis.msu.ru/content/mgu-2014-k-150-letiyu-moskovskogo-matematicheskogo-obshchestva.

[15] *Токарева Т. А.* Московское математическое общество после реорганизации и во время Великой Отечественной войны // Историко-математические исследования. Вторая серия. 2014. Вып. 15(50). С. 11–16.

[16] *Токарева Т. А.* Первые съезды отечественных математиков: предыстория и формирование Советской математической школы // Историко-математические исследования. Вторая серия. 2001. Вып. 6(41). С. 213–231.

[17] *Токарева Т. А.* Филоматический пролог Московского математического общества // Историко-математические исследования. Вторая серия. 2002. Вып. 7(42). С. 39–62.

[18] Устав Московского Математического Общества и извлечение из его протоколов // Матем. сб. 1867. Т. II, вып. 1. С. III–VIII.

[19] *Форд Ч.* Дмитрий Федорович Егоров: материалы из архива Московского университета // Историко-математические исследования. Вторая серия. Вып. 1(36), № 2. М., 1996. С. 146–164.

[20] *Юшкевич А. П.* История математики в России до 1917 года. М.: Наука, 1968.

С. С. Демидов, ИИЕТ РАН
serd42@mail.ru

С. С. Петрова, мехмат МГУ
serd42@mail.ru

В. М. Тихомиров, мехмат МГУ
vmtikh@googlemail.com

Т. А. Токарева, ИИЕТ РАН
tantokareva@yandex.ru

Александр Михайлович Абрамов (5 июня 1946 г. – 24 мая 2015 г.)

В. М. Тихомиров

Александр Михайлович Абрамов родился 5 июня 1946 г. в Астрахани.

Его отец, Михаил Григорьевич Абрамов (1917–1979) родился в Астрахани, закончил Ленинградский политехнический институт в 1941 году, до ноября 1941 года работал по профессии инженером-технологом, затем был призван в армию, участвовал в Великой Отечественной войне и служил в армии до декабря 1945 года. Демобилизовавшись, вернулся в Астрахань, где занимался активной общественной деятельностью (в частности, был заместителем председателя исполкома Астраханского областного совета депутатов трудящихся). Михаил Григорьевич был высокообразованным человеком, всю жизнь любил и собирал книги, вместе с женой активно участвовал в клубе книголюбов, своим примером воспитывая детей и внуков.

На долю матери А. М. Абрамова — Нины Михайловны Любовцовой (1912–2004) — выпала нелёгкая судьба: с двух лет росла без отца и матери. Окончив школу с педагогическим уклоном, работала в школе, потом стала студенткой Ленинградского технологического института. После института много лет работала по профессии, но мечтала о литературном труде. В семидесятые годы, выйдя на пенсию, целиком посвятила себя писательству. Её первая историческая повесть «Посольство в Каракорум», выпущенная в 1981 году, посвящена сопряжённому с очень большими опасностями походу Александра Невского в Каракорум — древнюю столицу монгольских

ханов — за ярлыком на княжение своими владениями. И другие исторические повести Нины Михайловны посвящены необычайно интересным событиям давно ушедших веков. До последних лет писательница жила насыщенной творческой жизнью.

А. М. Абрамов был очень тесно связан со своими родителями, гордился ими, заботился о них, любил родной город и постоянно навещал места, где прошло его детство.

Он учился в известной школе Астрахани — школе № 10. Его учительницей математики была Маргарита Александровна Зороастрова — замечательный педагог. Она увлекла своего ученика математикой. Контакт со своей учительницей А. М. сохранил на всю жизнь. В те годы складывалось шефство педагогических институтов над школами. Сотрудник астраханского педагогического института Юрий Алексеевич Изосимов открыл для многих ребят дорогу на областные и российские олимпиады.

Вся жизнь Александра Михайловича оказалась связанной с А. Н. Колмогоровым. Учась в седьмом классе, Саша заочно познакомился с Андреем Николаевичем, когда прочитал его брошюру «О профессии математика». Весной 1963 года Саша стал призёром Всероссийской математической олимпиады. Призы вручал А. Н. Колмогоров. Тогда произошла первая встреча Саши со своим будущим учителем.

Летом 1963 г. Саша принял участие в первой колмогоровской летней школе, проходившей в Красновидове. Красновидовская школа сыграла огромную роль во всей жизни Саши. Там произошло его знакомство со многими математиками — преподавателями школы: В. И. Арнольдом, А. А. Егоровым, А. А. Карацубой, А. Г. Кушниренко, А. Л. Тоомом и др., составившими первый круг его старших друзей. Впоследствии этот круг необычайно расширился, включив в себя множество людей разных поколений, профессий и жизненных интересов.

По окончании летней школы Абрамов был отобран для обучения в только что созданной физико-математической школе (ФМШ) № 18, будущем колмогоровском Интернате. И в летней школе, и в первые годы работы ФМШ огромную роль играл Андрей Николаевич Колмогоров как преподаватель, воспитатель и организатор. Впоследствии на протяжении всех последующих лет Андрей Николаевич испытывал особые чувства к первым девятнадцати выпускникам Интерната. И в самом деле, первый выпуск Интерната был замечательным, а А. М. Абрамов стал «историографом» своего интернатского класса, председателем Клуба выпускников ФМШ.

Вся последующая жизнь Александра Михайловича Абрамова также оказалась теснейшим образом связанной с колмогоровским Интернатом. По окончании ФМШ в 1964 году Абрамов поступает на мехмат МГУ

и уже в студенческие годы (с 1966 года) начинает преподавание в ФМШ. В студенческие годы Абрамов получил интересный результат по топологии, составивший его первую публикацию. Постепенно Саша вошёл в обширный круг научной школы Колмогорова. В те годы он много встречался с Колмогоровым, и на распределении Андрей Николаевич предложил Абрамову поступить к нему в аспирантуру. Так Александр Михайлович стал сподвижником Колмогорова в осуществлении его педагогических замыслов.

Тема диссертации Абрамова была связана с актуальной проблемой педагогики — построением школьного курса геометрии, обладающего математической завершённостью. Предложенная Колмогоровым аксиоматическая система, представлявшая развитие системы аксиом Евклида и вобравшая в себя некоторые черты математики нового времени, обладала одним (как написано в самом учебнике) «существенным пробелом», состоящим в недостаточно отчётливо определённом понятии «величина угла». И основной математической задачей, поставленной Колмогоровым перед своим аспирантом (помимо многих исторических и методических проблем), было построение теории измерения углов. Абрамов справился с поставленной задачей. Написание диссертации заняло свыше пяти лет. Она была защищена в 1976 году. Сейчас было бы важным издание этой диссертации, или, может быть, отдельное комментированное издание второй главы диссертации «Построение планиметрии на основе аксиом А. Н. Колмогорова».

После аспирантуры Александр Михайлович работал в Академии педагогических наук СССР, в издательстве «Просвещение» и во временном научно-исследовательском коллективе «Школа».

Но с самого начала, с 1970 года, Александр Михайлович стал принимать участие в работе над двумя основными учебниками, которые готовились под редакцией А. Н. Колмогорова. Учебник «Алгебра и начала анализа», написанный при участии Абрамова, был переиздан 24 раза. Первые учебные пособия по геометрии стали появляться в начале семидесятых годов, а окончательную форму учебник по геометрии приобрёл к концу 1970-х. В доработке этого учебника Александр Михайлович также принял активнейшее участие.

В восьмидесятые годы Александр Михайлович начинает многостороннюю организаторскую и общественную деятельность в области просвещения. Он был руководителем советской команды на Международных математических олимпиадах (1981–1984), советником министра образования СССР (1990–1991).

При активном участии А. М. Абрамова был организован МИРОС (Московский институт развития образовательных систем), и в 1992 году он возглавил коллектив института, в котором проработал до 2002 года. Эта

деятельность А. М. Абрамова заслуживает подробного освещения. В некотором отношении девяностые годы — это звёздный период его деятельности. А. М. Абрамов был избран членом-корреспондентом РАО (Российской академии образования), с 1997 г. по 2002 г. был членом Президиума РАО.

К концу этого периода А. М. опубликовал свыше двухсот статей, посвящённых истории и работе ФМШ, проблемам образования, творчеству Колмогорова и другим вопросам.

Последние пятнадцать лет Александр Михайлович посвятил обширной и разносторонней публицистической деятельности, посвящённой школьному (в основном математическому) образованию, истории колмогоровского Интерната, разборке и изданию колмогоровского архива и множеству других проблем просвещения.

Александр Михайлович был человеком очень высокой культуры. Его гуманитарные интересы были необычайно широки — поэзия, литература, мировая история, история науки и многое другое. Во время его директорства в МИРОСе было издано множество выдающихся книг, учебников, учебных пособий. В качестве примеров можно привести первую книгу, изданную в МИРОСе, — «Евангелие и древнерусская литература» Н. В. Давыдовой, книгу литературоведа, писателя и мемуариста В. Я. Лакшина «Берега культуры» о сохранении и развитии отечественной культуры, предназначенную преподавателям-словесникам и учащимся старших классов, книгу о Гоголе Юрия Манна, переводы Сергея Аверинцева. Можно назвать такие книги, как «Час души» для детского чтения — сборник прекрасных произведений русских писателей XIX–XX вв., «Век Наполеона: люди и судьбы», «Мифологический словарь школьника», книга о московских гимназиях и множество других интересных необычных книг.

Александр Михайлович был счастлив и любим в своём семейном кругу. Вместе со своей женой Натальей Григорьевной Пучковой, выпускницей химфака, он создал замечательный, прекрасно обставленный Дом, с очень богатой библиотекой, где в кабинете хозяина дома сохраняется большой хорошо структурированный архив. Московский университет был их местом встречи и навсегда остался alma-mater и центром притяжения их обоих.

При разрешении многих жизненных ситуаций, когда возникала гамлетовская альтернатива (из монолога «Быть или не быть»): согнуться под жестокими ударами судьбы или, в противостоянии с морем бедствий, покончить с ними, Александр Михайлович Абрамов неизменно избирал второй путь. Он был истинным часовым на рубеже просвещения. И смерть его тяжёлой ношей легла на плечи всех, кто его знал.

Жизнь Александра Михайловича оборвалась на взлёте его творчества. Он имел чрезвычайно большие замыслы. На очереди дня стояло опубли-

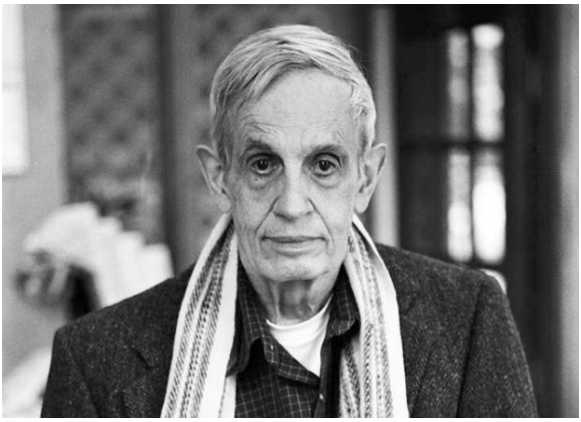
кование почти полностью подготовленного им к изданию полного списка трудов и энциклопедических статей Колмогорова. А. М. принимал активное участие в подготовке к изданию публицистических работ Колмогорова и его работ, посвящённых проблемам образования. Завершение этих его замыслов, как можно надеяться, произойдёт в недалёком будущем. Но перед друзьями и коллегами А. М. стоит задача разборки его архива, издание сборника воспоминаний о нём и подготовка к изданию его трудов.

В. М. Тихомиров, мехмат МГУ
vmtikh@googlemail.com

Торг с богами

В. А. Гурвич

«Кого боги хотят погубить,
того они лишают разума»,
Верно ли обратное?



Джон Форбс Нэш младший — единственный в мире обладатель и Нобелевской, и Абелевской премий. Получив последнюю, он с супругой Алисией возвращался из Норвегии домой в Нью-Джерси 23 мая 2015 г. Они взяли такси в аэропорту Ньюарк и... оба погибли в автокатастрофе. Впрочем, было бы удивительно, если бы человек с такой биографией спокойно умер в своей постели. Дело в том, что боги и в самом деле лишили его разума, в прямом смысле слова. Не менее тридцати лет Джон Нэш страдал параноидальной шизофренией в крайне тяжёлой форме. Удалось ли его этим погубить? Во всяком случае, он боролся на равных: победил неизлечимую болезнь, получил две высшие награды: за вклад в экономику и в математику. Возможно, чаша терпения богов переполнилась, они отказались от изощрённых методов наказания и прибегли к вульгарным, но более надёжным. Капитулировали?

Ранние годы и обучение. Джон Нэш родился 13 июня 1928 г. в штате Западная Вирджиния, в маленьком городке Блуфилд в горах Аппалачи. Отец, Джон Форбс Нэш старший, — инженер-электрик; мать, Маргарита Вирджиния, до замужества 10 лет преподавала в школе. Она занималась с сыном дома, так что он часто пропускал занятия в средней школе. В своей автобиографии [51] он пишет, как в 14 лет прочитал книгу «Творцы математики» Эрика Белла и сумел без посторонней помощи доказать малую теорему Ферма. В последнем классе он усиленно занимался математикой, посещал курсы в местном колледже. После школы Нэш поступил в Технологический институт Карнеги (теперь университет Карнеги — Меллона) в Питтсбурге, получив стипендию Джорджа Вестингауза (таких всего 10 на всю страну). Сначала он специализировался в инженерной химии, потом вообще в химии, прослушал курс международной экономики и, наконец, по совету своего преподавателя Джона Лайтона Синга переключился на математику. В 1948 г. он получил степень бакалавра, затем магистра и поступил в аспирантуру Принстона. Его научный руководитель, Ричард Даффин, написал ему весьма лаконичную рекомендацию: «Он — математический гений». Нэш получил предложение и из Гарварда, однако выбрал Принстон, где руководитель математического отделения Соломон Лефшец назначил ему именную стипендию Джона С. Кеннеди. Сыграло роль и то, что Принстон намного ближе к Блуфилду, чем Гарвард.

На Нэша произвела сильное впечатление вышедшая в 1944 г. книга Джона фон Неймана и Оскара Моргенштерна «Теория игр и экономическое поведение» [69]. Она натолкнула Нэша на великолепную идею — ввести понятие равновесия для игр n лиц. Оно называется теперь равновесием Нэша (РН) и непосредственно обобщает на случай произвольного числа игроков классическое понятие *седловой точки* для игр двух лиц с нулевой суммой, так называемых матричных игр. Ему удалось обобщить теорему фон Неймана (1928) о существовании седловой точки в смешанных стратегиях для конечных матричных игр и доказать существование РН в смешанных стратегиях для конечных игр n лиц в нормальной форме. Теорема Нэша (1950) оказалась простым следствием не так давно появившейся теоремы Какутани (1941) о неподвижной точке. Это основной результат диссертации Нэша, которую он защитил в Принстоне в 1950 г. Ему ещё не исполнилось и 22 лет; текст уложился в 30 страниц.

Ниже мы дадим краткое математическое введение, включающее все три теоремы, а затем вернёмся к биографии.

Равновесие Нэша и седловая точка. Пусть $I = \{1, \ldots, n\}$ — множество игроков, а X_i — множество стратегий игрока $i \in I$. Рассмотрим прямое произведение $X = X_1 \times \ldots \times X_n$ этих n множеств. Его элементы

$x = (x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n = X$ называются *профилями стратегий* или *ситуациями. Игра n лиц в нормальной форме* определяется как отображение $u\colon X \to \mathbb{R}^n$, которое каждому профилю стратегий $x \in X$ ставит в соответствие n-мерный вещественный вектор $u(x) \in \mathbb{R}^n$ — *профиль выигрышей*, чья координата $u_i(x)$ интерпретируется как *выигрыш* или *платёж игрока $i \in I$ в ситуации $x = (x_1, \ldots, x_n) \in X$*.

Ситуация x называется *равновесием Нэша* (РН), если $u_i(x) \geqslant u_i(x')$ для любого игрока $i \in I$ и для любой ситуации $x' \in X$, совпадающей с x в каждой координате, кроме, быть может, i. Иными словами, ни один из игроков не может, выбрав новую стратегию, увеличить свой выигрыш при условии, что все остальные придерживаются своих старых стратегий.

Заметим, что коалиция, состоящая более чем из одного игрока, вообще говоря, может увеличить выигрыш всех своих участников при выборе ими новых стратегий, даже если все остальные игроки придерживаются старых. По этой причине РН относится к *бескоалиционным* или *некооперативным* концепциям решения.

Игра двух лиц с нулевой суммой (в нормальной форме) определяется условиями: $n = 2$ и $u_1(x) + u_2(x) = 0$ в любой ситуации $x \in X$. Такая игра однозначно определяется вещественной таблицей $u_1\colon X_1 \times X_2 \to \mathbb{R}$ и поэтому называется *матричной игрой*. Игрок 1 максимизирует u_1, а игрок 2 минимизирует, поскольку он максимизирует u_2.

РН $x = (x_1, x_2)$ в матричной игре называется седловой точкой (функции u_1). Такое название объясняется тем, что график функции u_1 на произведении $X = X_1 \times X_2$ имеет форму седла: u_1 достигает минимума в строке x_1 и максимума в столбце x_2:

$$u_1(x_1', x_2) \leqslant u_1(x_1, x_2) \leqslant u_1(x_1, x_2') \quad \text{для всех } x_1' \in X_1 \text{ и } x_2' \in X_2.$$

Пусть $x = (x_1, x_2)$ и $x' = (x_1', x_2')$ — две седловые точки. По определению, имеем цепочку неравенств:

$$u_1(x_1', x_2) \leqslant u_1(x_1, x_2) \leqslant u_1(x_1, x_2') \leqslant u_1(x_1', x_2') \leqslant u_1(x_1', x_2).$$

Отсюда сразу следует, что все четыре числа равны:

$$u_1(x_1, x_2) = u_1(x_1, x_2') = u_1(x_1', x_2') = u_1(x_1', x_2)$$

и что две появившиеся ситуации (x_1, x_2') и (x_1, x_2') также являются седловыми точками.

Таким образом, хотя седловых точек может быть несколько, но

(i) седловое значение единственно (оно называется *значением игры*) и

(ii) все седловые точки образуют прямоугольный массив

$$X^* = X_1^* \times X_2^* \subseteq X_1 \times X_2 = X.$$

При этом любые стратегии $x_1 \in X_1^* \subseteq X_1$ и $x_2 \in X_2^* \subseteq X_2$ называются *оптимальными стратегиями* игроков 1 и 2 соответственно. Свойство (ii) очень важно; оно означает, что игрокам не нужно договариваться, достаточно просто каждому применить оптимальную стратегию.

Игра двух лиц в нормальной форме называется биматричной. К сожалению, РН может не обладать перечисленными выше приятными свойствами уже в биматричной игре, если она имеет ненулевую сумму. Биматричная игра может иметь несколько РН, в которых игроки могут иметь разные выигрыши. При этом если игроки применяют «оптимальные» стратегии, соответствующие разным РН, то полученная ситуация может вообще не быть РН. Более того, РН не обязано даже быть Парето-оптимумом; иными словами, иногда оба игрока могут увеличить свои выигрыши, применив новые стратегии одновременно. При этом они образуют коалицию, а как уже упоминалось, РН — бескоалиционная концепция, гарантирующая устойчивость только относительно индивидуальных действий игроков.

Рассмотрим таблицы ниже, дающие два примера биматричных игр 2×2. Два игрока, 1 и 2, выбирают строки и столбцы биматрицы; в каждой её клетке (ситуации) их выигрыши расположены в левом нижнем и правом верхнем углу и сравниваются они в столбцах и строках соответственно. Разумеется, биматрицу можно представить как две матрицы с выигрышами игроков 1 и 2.

Семейный спор

1 2	0 0
0 0	2 1

Дилемма заключённого

-1 -1	0 -3
-3 0	-2 -2

Первый пример называется «семейный спор». Пара намерена провести вечер либо в ресторане, либо на футболе. Дама, игрок 1, предпочитает ресторан, а господин, игрок 2, — футбол. Если каждый выбирает свою первую стратегию, то они идут в ресторан (выигрыши равны 2 и 1), если вторую, то на футбол (выигрыши равны 1 и 2); если же они выбрали стратегии с разными номерами, 1 и 2 или 2 и 1, то остаются дома (с выигрышем 0 каждый). Нетрудно проверить, что в игре два РН с разными векторами выигрышей $(2, 1)$ и $(1, 2)$, а также что эти РН не заполняют

прямоугольный массив. Таким образом, понятие оптимальной стратегии вообще не определено.

Второй пример называется «дилемма заключённого». Два сообщника совершили какое-то довольно серьёзное правонарушение и попали под подозрение. У обвинителя, однако, проблемы с доказательствами, и он предлагает каждому из них сделку: сознаться и выйти на свободу, но уж тогда товарища упекут «по полной», скажем на 3 года. Соответствующие две ситуации в биматрице находятся на побочной диагонали; выигрыши в них равны $(0, -3)$ и $(-3, 0)$ соответственно. Каждый выигрыш равен числу лет в тюрьме, взятому, естественно, со знаком минус. Если оба «идут в несознанку», то получают по одному году (скажем, за неуплату налогов; выигрыши $(-1, -1)$), а если оба сознаются, то по 2 года (с учётом раскаяния и сотрудничества; выигрыши $(-2, -2)$). Легко видеть, что $(-2, -2)$ — единственное РН. Однако эта ситуация не оптимальна по Парето. Вступив в коалицию, игроки могут получить $(-1, -1)$, но эта ситуация не равновесна, поскольку каждый может улучшить результат, предав товарища. Моральные аспекты теория игнорирует, принимаются в расчёт лишь выигрыши.

Когда матричные игры решаются в чистых стратегиях. Пока что речь шла о проблемах, возникающих, когда имеется несколько РН, а что делать, если их вообще нет? Такое вполне возможно уже для матричных игр. Читатель легко проверит, что матричная игра 2×2 не имеет седловой точки тогда и только тогда, когда «одна диагональ больше другой»; точнее, каждый из элементов одной из диагоналей строго больше каждого из элементов другой. Следующий общий критерий имеет место для любой матричной игры. Рассмотрим числа

$$v_1 = \max_{x_1 \in X_1} \min_{x_2 \in X_2} u_1(x_1, x_2), \qquad v_2 = \min_{x_2 \in X_2} \max_{x_1 \in X_1} u_1(x_1, x_2),$$

которые называются *максимин* и *минимакс*. Они интерпретируются как наилучшие результаты, которые игроки 1 и 2 могут *гарантировать* при любых действиях противника. В самом деле, легко видеть, что игрок 1, максимизирующий u_1, может гарантировать v_1, но не больше, а игрок 2, минимизирующий u_1, может гарантировать v_2, но не меньше. Соответствующие стратегии игроков называются осторожными или безопасными. Читатель без труда докажет, что неравенство $v_1 \leqslant v_2$ имеет место в любой матричной игре. Оно и понятно: гарантировать что-либо непросто. Более того, $v_1 = v_2$ тогда и только тогда, когда седловая точка существует. При этом она является РН и число $v = v_1 = v_2$ является *значением* игры.

Это значение — единственный претендент на «высокое звание оптимума». В самом деле, игроку 1 не стоит надеяться на результат $> v = v_1$, по-

скольку у противника есть стратегия, гарантирующая $\leqslant v_2 = v$; аналогично игроку 2 не стоит рассчитывать на результат $< v = v_2$, поскольку у противника есть стратегия, гарантирующая $\geqslant v_1 = v$. Таким образом, любая пара осторожных стратегий игроков 1 и 2 является оптимальной и реализует седловую точку (РН, другими словами). К сожалению, такая простота и определённость имеют место только для игр двух лиц с нулевой суммой.

Замечание. Даже если седловая точка существует, её не всегда просто найти. Например, шахматы вполне возможно представить как матричную игру с выигрышами, принимающими всего три значения: $\{\pm 1, 0\}$ — выигрыш белых, выигрыш чёрных и ничья. В 1912 г. Эрнст Цермело доказал, что в случае шахмат $v_1 = v_2$ и, следовательно, седловая точка существует [70]. Однако матрица шахматной игры имеет такие огромные размеры, что ни оптимальных стратегий, ни оптимального результата (значения игры) никто не знает, хотя они и существуют. Возможно, оптимальный результат — ничья, а может, белые начинают и выигрывают, а возможно, и чёрные, этого тоже нельзя исключить. На практике выигрывает тот, кто лучше играет. «Мы видим, что блондин играет хорошо, а брюнет играет плохо. И никакие теории игр не изменят этого соотношения сил», — как мог бы сказать Остап Бендер.

Теорема фон Неймана о решении матричных игр в смешанных стратегиях. Что же делать, когда $v_1 < v_2$ и, значит, седловых точек нет? Допустим, игроки разыгрывают данную матричную игру не один раз, а целый вечер, то есть десятки или даже сотни раз. Если игрок i применяет всё время безопасную стратегию, то оппонент $3 - i$ это заметит и ограничит его гарантированным результатом v_i, где $i = 1, 2$. Это не очень-то вдохновляет. Ясно, что оптимальный результат лежит где-то в промежутке $[v_1, v_2]$ и, чтобы получить его, надо время от времени рисковать. Каждый игрок в покер знает, что блефовать, хотя бы изредка, необходимо. Но когда именно и как часто? Можно пытаться применить психологию, например усыпить бдительность противника, играя осторожно, а потом «вдруг» рискнуть. Но противник — тоже психолог... К сожалению, на этом пути не удаётся получить хорошей теории. Её даёт математический подход, а именно идея случайного выбора. Лучший способ скрыть от противника свой следующий ход — не знать его самому.

Итак, пусть дана матричная игра $u_1 \colon X_1 \times X_2 \to \mathbb{R}$ с $k_1 \times k_2$-матрицей, то есть $|X_1| = k_1$ и $|X_2| = k_2$. Пусть Y_i обозначает множество всех вероятностных распределений на множестве стратегий X_i игрока $i = 1, 2$. Эти распределения $y_i \in Y_i$ называются *смешанными* стратегиями, а $x_i \in X_i$ будут отныне называться *чистыми* стратегиями игрока i. Таким образом,

мы расширяем исходную конечную матричную игру $u_1\colon X_1 \times X_2 \to \mathbb{R}$ в чистых стратегиях, получая бесконечную «матричную» игру $U_1\colon Y_1 \times Y_2 \to \mathbb{R}$ в смешанных стратегиях. Предполагая, что игроки $i = 1, 2$ выбирают свои чистые стратегии $x_i \in X_i$ независимо и в соответствии с распределениями $y_i \in Y_i$, получаем

$$U_1(y_1, y_2) = \sum_{\substack{x_1 \in X_1,\\ x_2 \in X_2}} u_1(x_1, x_2) y_1(x_1) y_2(x_2),$$

где $y_i(x_i)$ — вероятность выбора чистой стратегии x_i игроком $i \in \{1, 2\}$, применяющим смешанную стратегию y_i. Таким образом, U_1 — билинейная форма, определённая $k_1 \times k_2$-матрицей исходной игры на произведении двух вероятностных симплексов в пространствах $\mathbb{R}^{k_1}$ и $\mathbb{R}^{k_2}$. По аналогии с v_1 и v_2 введём числа

$$V_1 = \max_{y_1 \in Y_1} \min_{y_2 \in Y_2} U_1(y_1, y_2), \qquad V_2 = \min_{y_2 \in Y_2} \max_{y_1 \in Y_1} U_1(y_1, y_2),$$

которые называются *максимин* и *минимакс* в смешанных стратегиях. Они по-прежнему интерпретируются как результаты, которые игроки 1 и 2 могут гарантировать в смешанных стратегиях при любых смешанных (или, что на самом деле в силу выпуклости равносильно, — чистых) стратегиях противника. Строго говоря, следовало бы заменить max на sup и min на inf, поскольку множества смешанных стратегий бесконечны. Однако все супремумы и инфимумы достигаются, поскольку функция $U_1\colon Y_1 \times Y_2 \to \mathbb{R}$ линейна по каждому аргументу на прямом произведении симплексов.

Благодаря этому неравенство $V_1 \leqslant V_2$ может быть доказано так же, как и $v_1 \leqslant v_2$. Кроме того, легко видеть, что $v_1 \leqslant V_1$, а $v_2 \geqslant V_2$. В самом деле, поскольку множество смешанных стратегий шире, чем множество чистых, то и гарантировать в них можно результат уж во всяком случае не хуже.

Итак, возникает цепочка неравенств $v_1 \leqslant V_1 \leqslant V_2 \leqslant v_2$. Главный вопрос: выполняется ли равенство $V_1 = V_2$ для любой матричной игры? Если так, то никакая психология не нужна и $V = V_1 = V_2$, значение игры в смешанных стратегиях — единственный претендент на оптимальность.

Однако это вовсе не казалось очевидным. В 1920-х годах этим вопросом занялся Эмиль Борель [17–19], создатель теории меры. Он рассматривает симметрические матричные игры, определяемые условиями $X_1 = X_2 = \mathbf{X}$ и $u_1(x_1, x_2) + u_1(x_2, x_1) = 0$ для любых $(x_1, x_2) \in \mathbf{X}$ (в частности, $u_1(x, x) = 0$ при всех $x \in \mathbf{X}$) или, иными словами, — заданные квадратными антисимметрическими матрицами. В симметрической игре значение, если существует, должно быть равно нулю, а оптимальные смешанные стратегии двух игроков, если существуют, могут быть выбраны равными. Однако

существование доказать совсем не просто; на самом деле, не проще, чем в общем случае, поскольку любая матричная игра может быть сведена к симметрической лишь немного большего размера. Такое замечание имеется в работе Бореля [17], но строго это было обосновано лишь в 1950-е годы Брауном и фон Нейманом. Борель доказывает, что $V_1 = V_2 = 0$ для симметрических игр размера 3×3 и 5×5, но сомневается, что результат верен в общем случае, считая, что помимо рандомизации тут понадобится и психология.

Равенство $V_1 = V_2$ было доказано в 1928 г. фон Нейманом [68]; ему было всего 25 лет. Доказательство довольно сложное; в статье целых 26 страниц.

В 1950-е годы была установлена эквивалентность симметрических матричных игр парам двойственных (совместных и ограниченных) задач линейного программирования. При этом $-V_2$ и $-V_1$ — оптимальные значения: максимум прямой и минимум двойственной задачи, а значит, $V_1 = V_2$ согласно сильной теореме двойственности. Этот замечательный результат был получен в 1950-х в работах Гейла — Куна — Такера, Брауна и Данцига. По свидетельству последнего, про связь между линейным программированием и матричными играми ему говорил фон Нейман ещё в 1947 г.

Теоремы Брауэра, Какутани и Нэша. В 1950 и 1951 гг. Нэш обобщил теорему фон Неймана для матричных игр на игры n лиц в нормальной форме. А именно, он доказал, что для последних тоже существует РН в смешанных стратегиях. Его доказательство опирается на топологические теоремы Брауэра и Какутани о неподвижной точке.

Первая утверждает, что любое непрерывное отображение выпуклого компакта Y (в евклидовом пространстве $\mathbb{R}^n$) в себя имеет неподвижную точку. Это одна из многочисленных версий теоремы Брауэра. Заметим, что и выпуклость, и замкнутость, и ограниченность Y, и непрерывность отображения — существенные условия. При отсутствии любого из них нетрудно построить контрпример. (Сделайте это!)

Теорема Какутани обобщает теорему Брауэра следующим образом. Пусть снова имеется выпуклый компакт $Y \subseteq \mathbb{R}^n$ и отображение $f\colon Y \to 2^Y$, которое ставит в соответствие каждой точке $y \in Y$ непустое выпуклое подмножество $f(y) \subseteq Y$. Тогда если график f замкнут, то имеется неподвижная точка $y \in f(y)$.

Говорят, что график отображения $f\colon Y \to Z$ замкнут, если множество всех пар $\{(y,z)\,|\,z \in f(y)\}$ замкнуто в топологии прямого произведения $Y \times Z$.

Отступление. Какутани работал над своей теоремой в Принстоне во время визита в США в 1941 г. В декабре между Японией и США началась война. Тем не менее, Какутани смог завершить визит и вернуться в Осаку летом 1942 г. Кен Бинмор в восьмой главе своей книги [16]

вспоминает, как Какутани подошёл к нему на одной из конференций и спросил, почему на его доклад пришло так много экономистов. «Вероятно, их интересовали приложения теоремы Какутани», — ответил Бинмор. «Какой теоремы?» — спросил Какутани. Похожий анекдот рассказал мне Стеф Тайс в 1990 г. На какой-то конференции докладчик упомянул одну из гипотез ван дер Вардена. Пожилой человек в первом ряду осведомился, что это за гипотеза. Докладчик ответил, что в двух словах не расскажешь, а время ограничено. Потом ему объяснили, что и пары слов хватило бы, поскольку спрашивал как раз ван дер Варден.

Теорема Нэша — несложное следствие теоремы Какутани. Пусть $Y = Y_1 \times \ldots \times Y_n$ — множество всех ситуаций в смешанных стратегиях некоторой игры n лиц. Функция $f(y)$ определяет все ситуации в Y, для которых стратегия каждого из n игроков является наилучшим ответом в ситуации y. Поскольку наилучший ответ не обязан быть единственным, $f(y)$ — это, вообще говоря, множество ситуаций. Нетрудно проверить, что все условия теоремы Какутани выполнены, а также что включение $y \in f(y)$ означает в точности, что y является равновесием Нэша. (Проверьте и то, и другое.)

Эта теорема появилась в статье Нэша [54] и в его диссертации [56] в 1950 г. В следующей статье [55] он выводит свой результат непосредственно из теоремы Брауэра, считая, что это значительное упрощение. Новое доказательство действительно коротко (см. § 2 в [55]), но и старое (план которого изложен в предыдущем абзаце) тоже трудным не назовёшь. Видимо, Нэш принимает в расчёт, что теорема Какутани сложнее теоремы Брауэра.

В том же 1950 г. Нэш опубликовал в «Эконометрике» ещё одну фундаментальную статью, по теории переговоров [59]. В 1951 г. он покинул Принстон и перебрался в Массачусетский технологический институт, где занялся «чистой» математикой: алгебраической геометрией и уравнениями в частных производных.

В 1994 г. Нэш[1)] получил Нобелевскую премию по экономике[2)] «за фундаментальный анализ равновесия в теории некооперативных игр», то есть именно за три работы, написанные в 1950 г.

Заслуженно ли? Мнения разделились. Аргументы противников понятны. Сама по себе теорема Какутани непроста, но теорема Нэша выводится из неё так:

(i) надо ввести (многозначное) отображение наилучших ответов $f\colon Y \to 2^Y$ на множестве ситуаций в смешанных стратегиях;

[1)] И ещё двое учёных: Дж. Харсаньи и Р. Зелтен.

[2)] Официальное название — Премия Шведского национального банка по экономическим наукам памяти Альфреда Нобеля.

(ii) проверить, что выполнены все условия теоремы Какутани;

(iii) проверить, что неподвижная точка Какутани и равновесие Нэша — одно и то же.

План этот, вне всяких сомнений, весьма элегантен, но слишком прост: его реализация в подробном изложении занимает $5 + 20 + 5 = 30$ строк. Маловато для Нобелевской премии. Конечно, само определение РН дорогого стоит, но и оно занимает 3 строки. В диссертации Нэша 30 страниц, но он любит мотивировать, да и вообще поговорить на отвлечённые темы.

В лагере противников награждения Нэша был даже такой авторитетный американский математик, как Алан Хофман. Я об этом узнал случайно и гораздо позже, когда премия уже была вручена. Вообще-то в Америке не принято критиковать коллег, особенно таких знаменитых, как Джон Нэш. Но бывают и исключения. В 1998 г. я делал на семинаре в Ратгерсе доклад о РН (только в чистых, а не в смешанных стратегиях). В конце слово взял Хофман. Он кратко суммировал пункты (i, ii, iii) и сказал, что премию, в сущности, дали за одну страницу. На этом он не остановился и обрушился на теорию игр вообще. Точнее, он сказал, что эта «красивая и цветущая ветвь теории оптимизации» всё время претендует на нечто большее: на приложения в экономике и даже в политике.

Неожиданно Хофман обратился ко мне за поддержкой: «Вам и докладчик подтвердит, что всё это — лишь красивая математика, а никаких особенных практических приложений РН, да и вообще теория игр, иметь не могут». Я задумался. Согласиться со знаменитым человеком, конечно, легко, но ведь потом придётся оправдываться перед коллегами. Я ответил уклончиво: «Почему же не могут? Вот я в 1994 г., после 25 лет занятий теорией игр, хотя и не получил Нобелевскую премию, но зато выиграл в лотерею грин-карту. Это ли не практическое приложение?!»

Три опоздания Нэша. У партии «противников» есть и ещё один очень сильный аргумент. Специальный случай РН был рассмотрен аж в 1838 г. Антуаном Огюстеном Курно в его теории олигополий [24]. Имеется n фирм, выпускающих одну и ту же продукцию. Они принимают решение об объёмах производства. Каждая фирма стремится максимизировать свою прибыль, зависящую, вообще говоря, от всех n решений, но, конечно, не произвольно, а некоторым специальным образом. Курно ввёл подобающие случаю ограничения и ту же концепцию равновесия, что и Нэш через 111 лет. Курно рассмотрел также ключевое понятие *наилучшего ответа* участника на действия остальных. По определению, в любой ситуации равновесия стратегия любого игрока является наилучшим ответом на данные стратегии остальных. Такой подход позволяет искать («нащупывать», как

он выразился) равновесие, используя динамические (итеративные) методы. Некоторые авторы справедливо называют РН равновесием Нэша — Курно.

Ещё раз Нэш опоздал в 1948 г., но уже не на 111 лет, а всего на 6. Обучаясь в аспирантуре, он переоткрыл «Гекс», замечательную «игру соединений» на гексагональной решётке. Игра стала популярной в Принстоне, где игроки называли её «Нэш» или «Джон». Говорят, они играли на кафельной плитке в туалете. Однако изобретена эта игра была в 1942 г. датским математиком Питом Хайном в Институте Нильса Бора в Копенгагене. Там её называли «Полигон». В 1952 г. фирма «Братья Паркеры» выпустила игру под именем «Гекс»; оно и прижилось.

И, наконец, самое обидное опоздание Нэша, всего на несколько месяцев, — с 19-й проблемой Гильберта: «Все ли решения регулярных проблем вариационного исчисления аналитические?». После нескольких важных частичных результатов (см. ниже) окончательный, и утвердительный, ответ на этот вопрос Давида Гильберта получили независимо Эннио де Джорджи (1956, 1957) [25, 26] и Джон Нэш (1957, 1958) [53, 57]; см. также [50]. Ходили слухи, что если бы последний шаг был сделан не в двух независимых работах, а в одной, то автор получил бы Филдсовскую медаль. Может быть... Кто знает? Так или иначе, Нэш получил за эти работы в 2015 г. премию Абеля, которая котируется никак не ниже — и уж во всяком случае намного дороже в денежном выражении.

Гарольд Вильям Кун — бесспорный лидер партии сторонников Нэша, человек очень известный и весьма замечательный. (Помните теорему Гейла — Куна — Такера и условия Каруша — Куна — Такера?) Альберт Вильям Такер был научным руководителем Нэша (а также Ллойда Шепли и Давида Гейла) в Принстоне. В своей статье [37] в 1950 г. Кун уже вовсю оперирует равновесием Нэша.

Я познакомился с Куном в Москве в конце 1980-х годов, когда он был гостем Центрального экономико-математического института. Кун рассказал много любопытного, в том числе и про Нэша.

РН Джон придумал в 1948 г., а в 1950 г. уже защитился; тема — «Некооперативные игры», руководитель — Альберт Вильям Такер. Сам Кун защитился в том же году по более традиционной теории групп, руководитель — Ральф Харцлер Фокс. Другим его аспирантом был Джон Уиллард Милнор, филдсовский лауреат 1962 г. и абелевский лауреат 2011 г.

В 1959 г. у Джона (вскоре после женитьбы, на волне творческих успехов, совершенно неожиданно даже для самых близких друзей) вдруг началось тяжелейшее умственное расстройство... и продолжалось «добрых» 30 лет. Весной 1959 г. Джона поместили на два месяца в одну из лучших клиник Бостона, где поставили диагноз «параноидальная шизофрения», но помочь

толком не смогли. Потом, когда деньги у Джона кончились, его лечили в Трентоне, применяя инсулиновую шоковую терапию, и это помогло, хотя уже тогда этот метод считался варварским. В результате галлюцинации на время исчезли, но и всё прочее тоже ушло из его головы. Тогда Кун ему выхлопотал офис на факультете математики, куда он приходил каждый день. Студенты прозвали его фантомом Файн-Холла (название здания). Лет через двадцать состояние Нэша улучшилось и сейчас (в конце 1980-х годов) почти в порядке, хотя, конечно, ни о каких лекциях не может быть и речи. Галлюцинации появились вновь, но Джон научился довольно ловко экзаменовать их на реальность.

Кун также рассказал, какая уникальная компания «игроков» собралась в Принстоне в 1940–1950-х. Какутани доказал там свою теорему о неподвижной точке, приехав в Принстон из Осаки; ему с трудом удалось выбраться обратно после атаки на Пёрл-Харбор и начавшейся войны. Это было ещё до прибытия Куна в Принстон. В 1947 г. Джордж Данциг рассказал Джону фон Нейману про линейное программирование и недавно открытый им принцип двойственности; тот среагировал тотчас и сказал, что это «всего лишь» теорема о матричных играх, доказанная им в юности; а когда Нэш сформулировал своё обобщение этой теоремы, фон Нейман, не дожидаясь никаких доказательств, сказал: «Ну, это тривиально! — теорема о неподвижной точке, в чистом виде». Он вообще отличался мгновенной реакцией и изрядной «прямотой выражений». Альберт Такер сразу же позвал Куна и Давида Гейла обобщать теорию двойственности с линейного на выпуклое программирование. Такер был научным руководителем и у Ллойда Шепли, тоже защитившегося в Принстоне в 1953 г. В своей диссертации «Аддитивные и неаддитивные функции множеств» Шепли заложил основы теории кооперативных игр, ввёл свой знаменитый вектор (вектор Шепли) и был в одном шаге от понятия ядра. Его ввёл в 1959 г. Дональд Брюс Джиллис [32], близкий друг Джона; он тоже защитился в Принстоне под руководством самого фон Неймана. Важные результаты о ядре получил в 1967 г. Герберт Эли Скарф [63], бывший в середине 1950-х годов аспирантом в Принстоне.

В 1953 г. Шепли открыл ещё и стохастические игры, раньше чем появились управляемые марковские процессы — случай одного игрока. Дин Жилет из Беркли доказал разрешимость стохастических игр двух лиц с нулевой суммой (и с нулевой вероятностью остановки) с помощью классической теоремы Харди — Литлвуда (1931). Он не сумел аккуратно проверить её условия, и эту «дыру» удалось залатать только в конце 1960-х годов Лигету и Липману [44].

Большую часть всех этих рассказов теперь легко найти в интернете, но тогда почти всё я слышал впервые. Впрочем, кое-что забавное в ин-

тернете не найти. Например, Куна ужасно злило, что его многие считают венгром. (Действительно, в Венгрии Кунов много.) Особенно он разозлился, когда на какой-то конференции кто-то спросил его: «А ваш коллега Кун — венгр?» «Мало „венгра“, так он ещё принял меня за Такера!» — возмущался Кун. «Ну, Вы сами отчасти виноваты; кто писал статьи про венгерский метод [39, 40]?» — парировал я.

Покончив с воспоминаниями, Кун вдруг спросил: «Как по-твоему, заслужил Джон Нобелевскую премию по экономике за РН?» Я вежливо (и, в общем-то, правдиво) ответил, что был бы очень рад этому, что большая часть моих работ по теории игр как раз про РН, но нипочём не дадут — за одно определение и одну теорему с доказательством на полстраницы! «Посмотрим, — сказал Кун. — Главное, мы оба считаем, что он её заслужил».

В начале 1993 г., когда я уже перебрался в Ратгерс, мечты Куна начали вдруг сбываться. Точнее, он начал осуществлять свои планы. С Джоном Милнором они составили список из 21 публикации Нэша и разослали его (в электронном виде) математикам и экономистам с просьбой дать отзыв. Я тоже это письмо получил и ответил, что РН — простая, но гениальная концепция, главная в теории некооперативных игр, и что именно её я изучаю почти во всех своих работах в этой области.

В октябре 1994 г. Нэш получил Нобелевскую премию по экономике «за анализ равновесия в теории некооперативных игр». Номинантов было пятеро: Роберт Ауман, Рейнхард Зельтен, Джон Нэш, Джон Харсаньи и Ллойд Шепли. Получили трое «средних»: Зельтен, Нэш и Харсаньи, а будь моя воля, я бы присудил (в первую очередь) как раз двум «крайним»: Ауману и Шепли. Впрочем, оба её тоже получили, но позже.

Как и положено нобелевскому лауреату, Нэш написал автобиографию [51]; также о нём написал его друг и будущий абелевский лауреат 2011 г. Джон Милнор [45, 46], будущий нобелевский лауреат 2007 г. Роджер Майерсон [48], целая группа коллег [41]; в 2001 г. вышел целый сборник о Нэше под редакцией Гарольда Куна и Сильвии Назар [42].

Леонид Витальевич Канторович занимался тематикой очень близкой к принстонской, и с его работ началось развитие математической экономики в СССР [7–15]. В 1938 г. к нему обратились сотрудники Центральной лаборатории Ленинградского фанерного треста с просьбой оптимизировать методы раскроя фанерного листа. Это событие сыграло важнейшую роль в развитии мировой прикладной математики. Уже в 1939 г. Канторович написал небольшую (всего 67 страниц) книжку [9] о приложениях линейного программирования в экономике. Эта работа не была замечена математиками, зато экономисты и партийные работники её вскоре заметили, но автору это ничего, кроме неприятностей, поначалу не принесло.

А времена были суровые. После того как Канторовичем был предложен оптимальный метод распила фанерного листа, этот метод попытались применить и к разрезанию стальных листов [13]. После внедрения методов оптимизации на производстве одной из фабрик инженерам удалось улучшить показатели, что привело, однако, к негативным последствиям: система социалистического планирования требовала, чтобы и в следующем году план был перевыполнен, что было принципиально невозможно при имеющихся ресурсах, поскольку найденное решение было абсолютным максимумом; кроме того, фабрика не выполнила план по металлолому, львиная доля которого складывалась из обрезков стальных листов. Руководство фабрики получило выговор и больше с математиками не связывалось.

Таким образом, чересчур эффективное использование ресурсов (металла) сыграло отрицательную роль. С другой стороны, линейное программирование показало, что достижение оптимальных результатов не всегда требует использования всех имеющихся ресурсов, как считалось в теории оптимизации ранее.

Про эту историю рассказывают ещё следующее. Поскольку показатели улучшились на 20 процентов, партийное начальство поздравило Канторовича и сообщило, что в следующем месяце ждут такого же прогресса. Он ответил, что это невозможно, поскольку его метод даёт оптимальный результат. Тогда ему напомнили про «саботаж в военное время» (дело было в 1942 г.). На другой день его вызвали снова и сказали, что погорячились, что всё понимают и что пусть будет не 20, а всего 10 процентов.

Надо отдать должное мужеству Канторовича, который настаивал на внедрении своего открытия [12]. Однако метод Канторовича был отвергнут как противоречащий Марксовой теории трудовой стоимости и заимствующий вместо этого положения буржуазных теорий. Но самые большие неприятности у Канторовича случились в разгар хрущёвской оттепели. В период работы в СО АН СССР в начале 1960-х годов после доноса партийных экономистов его поместили в психиатрическую лечебницу и только вмешательство брата — известного психиатра — помогло ему вскоре покинуть больницу.

После этого всё резко изменилось к лучшему. Развитие тех же идей в Принстоне убедило начальство, что учёный прав. (Давно замечено, что начальство в России любит поговорить про «отечественный приоритет», но приоритетные идеи тормозит, осуждая за «западные заимствования», а продвигает их только после того, как они начинают бурно развиваться на том же Западе.) В 1964 г. Канторович был избран академиком АН СССР по отделению математики; в 1965 г. за разработку метода линейного программирования и экономических моделей удостоен Ленинской премии.

В 1975 г. стал лауреатом Нобелевской премии по экономике (совместно с Тьяллингом Купмансом) за вклад в теорию оптимального распределения ресурсов. В Штатах многие считают, что вместо Канторовича премию должен был получить Данциг. А по-моему, правильно было бы не «вместо», а «вместе».

Джон Форбс Нэш младший в фильме «Игры разума». Кун совершенно не удовлетворился блестящим завершением нобелевской эпопеи. В 1998 г. Сильвия Назар, профессор Колумбийского университета в Нью-Йорке и, наверное, лучшая американская журналистка, пишущая о математике и экономике, опубликовала биографическую книгу «Прекрасный ум. Жизнь математического гения и нобелевского лауреата Джона Нэша» [49]. Я уверен, что Кун был главным консультантом, а скорее всего, и вдохновителем этого труда. Имя Сильвии Назар он упоминал ещё в Москве. Книга стала бестселлером и была переведена на 30 языков; она получила премию Национальной гильдии литературных критиков как лучшая биография и была номинирована на Пулитцеровскую премию. В 2001 г. по этой книге Рон Ховард снял фильм «Игры разума» с Расселом Кроу в главной роли.

Важная роль и у друга и соперника Нэша, прототипом которого был Кун. Они играют партию в го, в конце которой совершенный разум Нэша вдруг разбивается вдребезги, — это центральная сцена фильма.

По ходу фильма зритель вдруг обнаруживает, что многие персонажи, к которым он успел привыкнуть, существуют только в воображении героя. Между тем тот тоже уже разобрался, что к чему, и пытается их там (в своём воображении) уничтожить, но они не сдаются и борются за своё существование. Это вам не вампиры и привидения, тут совсем другое. Нэш — убийца галлюцинаций. Такого в кино, кажется, ещё не было. Особенно мне понравилась сцена, где Нэш проверяет реальность представителя Нобелевского комитета, приехавшего в 1994 г. в Принстон спросить его согласия на номинацию. Нэш подзывает студентку, в существовании которой не сомневается, и просит её подтвердить реальность нобелевского гонца.

Нэш справился с острой стадией заболевания, но после этого никаких лекций студентам читать, разумеется, не мог. Он посещал свой офис в Принстоне, и у него даже был аспирант, Сет Патинкин, защитивший в 2003 г. диссертацию «Локализация РН в играх многих лиц»; но у Сета был и второй руководитель — Яков Григорьевич Синай.

В 1994 г. после получения Нобелевской премии Нэшу не предложили прочитать традиционную часовую лауреатскую лекцию в Стокгольме, так как организаторы опасались за его состояние. Вместо этого он прочитал лекцию в Упсале; она была посвящена космологии. Он пытался обосновать уравнения, описывающие нерасширяющуюся вселенную.

В 2015 г. Нэш получил Абелевскую премию «За яркий и оригинальный вклад в теорию нелинейных дифференциальных уравнений в частных производных и её приложения к геометрическому анализу». Лауреатскую лекцию он прочитал 20-го мая, за три дня до гибели. Про космологию там речи нет, но общая теория относительности Эйнштейна и бозон Хиггса упоминаются часто.

Кроме Нобелевской и Абелевской премий, Нэш получил ещё премию фон Неймана на своё 50-летие в 1978 г. (тогда о лекциях не могло быть и речи) и премию Леруа Стила Американского матобщества на 70-летие, а точнее — в 1999 г.

Начиная со второй половины 1990-х годов его стали часто приглашать на конференции с пленарными докладами. 13 июня 2008 г. Нэшу исполнилось 80 лет. Юбилейные конференции были устроены в Принстоне и в университете Стоуни-Брук. После этого он отправился в Европу. В Голландии в качестве юбилейного подарка ему преподнесли довольно крупный бриллиант. Но как подарить его мужчине? Организаторы не растерялись, наверное, вспомнили «Игры разума» и вмонтировали бриллиант в торец шариковой ручки. (В конце фильма «Игры разума» принстонские профессора отдают герою свои шариковые ручки в знак признания его заслуг и превосходства. Это, разумеется, — чисто голливудский приём, выдумка, но голландский бриллиант — реальность.) Из Голландии Нэш заехал на Международную конференцию по теории игр и менеджменту в Петербурге, где сделал часовой доклад «Идеальные и асимптотически идеальные деньги союзников».

Надо отдать должное Нэшу, с годами он делал доклады всё лучше и лучше. Очевидно, болезнь отступала. Наоборот, Ллойд Шепли выступал на конференциях всё хуже и хуже. Он не болен, просто стар. Видно, как трудно ему сконцентрироваться. Где-то в начале 2000-х годов они сравнялись. Я слышал несколько их докладов. Оба говорили очень медленно, надолго останавливаясь после каждой фразы; иногда совсем теряли мысль и не могли продолжать. Конечно, их участие помогало организаторам поднять престиж конференции. Слушатели (зрители) рассматривали их как живые символы теории игр; щёлкали фотоаппараты, жужжали кинокамеры...

Мне кажется, куда уместнее приглашать таких людей в качестве юбиляров или даже просто «свадебных генералов»; в этом нет ничего унизительного. Прочитать Нобелевскую (или Абелевскую) лекцию — куда ни шло: это дань многолетней традиции. Шепли читал её в 2012 г., ровно через полвека после выхода работы [30] про устойчивые паросочетания, за которую премия была присуждена. Читал он по бумажке всего семь минут. Его соавтор Давид Гейл не дожил четырёх лет.

Джон Форбс Нэш младший в фильме «Блестящее безумие». Документальный фильм о Нэше «Американская история: Блестящее безумие» (2002) просто рассказывает, как всё было на самом деле, поэтому он куда жёстче, чем «Игры разума».

В фильме родственники и друзья, а также коллеги-математики рассказывают о Нэше. Очень советую посмотреть.

Всякий раз, когда лекарства начинали действовать и наступало значительное улучшение, Нэш переставал их принимать. Он сожалел об изменении своего состояния и называл периоды ремиссии насильственным возвращением к здравому смыслу, говорил, что несколько великих математических идей явились ему из Космоса вместе с галлюцинациями, а рациональное мышление ослабляет связь с Космосом, что лечение и возврат к нормальности — конформизм, в то время как безумие — тоже возможный выход. (Об этом писал и Пушкин[3)], но тут — бесценное свидетельство человека, знающего предмет не понаслышке.)

Только через 30 лет Нэш принял волевое решение избавиться от галлюцинаций, и этот уникальный эксперимент удался, как ни трудно в это поверить. Он перестал прислушиваться к голосам и стал «экзаменовать галлюцинации на реальность».

Его постепенное излечение приписывали изобретению чудодейственных лекарств, но он перестал принимать лекарства ещё в 1960-х годах. Конечно, заменой лекарствам стали жена Алисия, друзья и коллеги, не бросавшие его в течение долгих тридцати лет болезни. Сыграло роль и то, что математики вообще терпимы к отклонениям и странностям и очень ценят одарённость, тем более гениальность.

Но, кажется, мы чересчур увлеклись кино. Пора вернуться к математике.

Разрешимость по Нэшу в чистых стратегиях. Итак, Нэш показал, что РН в смешанных стратегиях существуют всегда, точнее, для любых конечных игр n лиц в нормальной форме. Разумеется, этот результат был распространён и на бесконечные игры, правда, при некоторых топологических ограничениях.

Впрочем, имеется два класса игр, для которых РН могут существовать уже в чистых стратегиях: стохастические и позиционные игры с полной информацией (ПИ)[4)]. Могут-то могут, но, как выяснилось, не всегда существуют. Рассмотрим этот вопрос подробнее.

[3)] «Не дай мне Бог сойти с ума», 1833.

[4)] Условие ПИ означает, что и позиция, и возможные ходы в ней, и их вероятности, если это позиция случая, известны всем участникам. Например, шахматы, шашки, го и нарды — игры с ПИ, а карточные игры и домино — нет.

Разрешимость игровых форм в чистых стратегиях. Для изучения игр в нормальной форме (особенно, матричных и биматричных игр) оказывается весьма удобным отделить структурные свойства от выигрышей. Пусть $I = \{1, \ldots, n\}$ — множество игроков, X_i — конечное множество стратегий игрока $i \in I$, далее, $X = X_1 \times \ldots \times X_n$ — множество ситуаций и $A = \{a_1, \ldots, a_p\}$ — множество исходов игры. Отображение $g\colon X \to A$ называется игровой формой, а $u\colon I \times A \to \mathbb{R}$ — функцией выигрыша (платёжной функцией); $u(i, a)$ — выигрыш игрока $i \in I$ в случае исхода $a \in A$. Пара (g, u) определяет игру n лиц в нормальной форме.

Игровая форма g называется *разрешимой по Нэшу*, если при любом выигрыше u игра (g, u) имеет хотя бы одно РН.

Игровая форма двух лиц g называется *антагонистически разрешимой*, если при любом выигрыше $u\colon \{1, 2\} \times A \to \mathbb{R}$ с нулевой суммой ($u(1, a) + u(2, a) = 0$ для любого исхода $a \in A$) игра (g, u) имеет седловую точку; при этом g называется ± 1-*разрешимой*, если дополнительно потребовать, чтобы функция u принимала только значения $+1$ и -1.

Игровая форма двух лиц $g\colon X_1 \times X_2 \to A$ называется *плотной*, если её строки и столбцы образуют пару двойственных гиперграфов (клатер и блокер) на множестве A. Это определение можно переформулировать многими способами, например, так. Рассмотрим функции $f_1\colon X_2 \to X_1$ и $f_2\colon X_1 \to X_2$, их графики $gr(f_1), gr(f_2) \subseteq X_1 \times X_2$ и их образы в A при отображении g, то есть $[f_1] = g(gr(f_1)), [f_2] = g(gr(f_2)) \subseteq A$.

Игровая форма $g\colon X_1 \times X_2 \to A$ плотна тогда и только тогда, когда $[f_1] \cap [f_2] \neq \varnothing$ при любых f_1 и f_2. Иными словами, выберем ответ игрока 1 на каждую стратегию игрока 2, и наоборот. Игровая форма g плотна тогда и только тогда, когда при любых таких выборках соответствующие два подмножества исходов из A пересекаются.

Следующие свойства игровых форм двух лиц эквивалентны:

(а) плотность,

(б) ± 1-разрешимость,

(в) антагонистическая разрешимость,

(г) разрешимость по Нэшу.

Эквивалентность (а), (б) и (в) была установлена Эдмондсом и Фалкерсоном в [27] и независимо в [1]. Впрочем, доказательство почти очевидно.

(б) $\Rightarrow$ (а). Пусть $[f_1] \cap [f_2] = \varnothing$. Положим $u(1, a) = 1$ при $a \in [f_1]$, $u(1, a) = -1$ при $a \in [f_2]$, $u(1, a) = 1$ или -1 (произвольно) при $a \notin [f_1] \cup [f_2]$ и, наконец, $u(2, a) = -u(1, a)$ при всех $a \in A$. Очевидно, при этом $-1 = v_1 < v_2 = 1$, значит, игра (g, u) не имеет седловой точки, а u — выигрыш с нулевой суммой, принимающий лишь значения ± 1; иными словами, g не является даже ± 1-разрешимой.

(а) $\Rightarrow$ (в). Пусть g антагонистически неразрешима, то есть существует выигрыш u с нулевой суммой, при котором игра (g, u) не имеет седловой точки. Как мы знаем, в этом случае $v_1 < v_2$. Выберем исход, реализующий минимум (соответственно, максимум) в каждой строке (соответственно, столбце). Очевидно, два полученных подмножества A не пересекаются. Стало быть, g не плотна.

Наконец, импликация (в) $\Rightarrow$ (б) очевидна из определений (б) и (в).

Свойство (г) было добавлено к списку {а, б, в} в [3], см. также [2]. Импликация (г) $\Rightarrow$ (в) очевидна из определений (г) и (в), но (а) $\Rightarrow$ (г) доказывается несколько сложнее, и мы не приводим здесь это доказательство.

Несколько примеров плотных и неплотных игровых форм двух лиц приведены ниже.

Плотные игровые формы

a_1	a_1
a_2	a_3

a_1	a_1	a_3	a_3
a_2	a_4	a_2	a_4

a_0	a_0	a_0
a_0	a_1	a_2
a_0	a_3	a_4

a_1	a_1	a_3
a_1	a_2	a_2
a_3	a_2	a_3

a_1	a_1	a_2
a_1	a_1	a_3
a_4	a_3	a_3

a_1	a_2	a_1	a_2
a_3	a_4	a_4	a_3
a_1	a_4	a_1	a_5
a_3	a_2	a_6	a_2

Неплотные игровые формы

a_1	a_2
a_2	a_1

a_0	a_1	a_2
a_0	a_2	a_1

a_1	a_1	a_2
a_4	a_0	a_2
a_4	a_3	a_3

О применении вышеприведённых результатов к анализу разрешимости по Нэшу позиционных игровых форм см. [20, 22, 33].

Определение плотности естественно расширяется на игровые формы n лиц при $n \geqslant 2$. Однако плотность при этом уже не является ни необходимым, ни достаточным условием разрешимости по Нэшу; примеры можно найти в [2, 3].

Замечание. Когда я в 1988 г. рассказал всё это Куну, он немедленно спросил, какова алгоритмическая сложность проверки плотности игровых форм (иными словами, двойственности гиперграфов). Я честно ответил, что не знаю. Сейчас вопрос Куна выглядит более чем естественно, но в 1980-х алгоритмическая сложность ещё не была в моде в Москве. Когда

я прибыл в Ратгерс в 1993 г., оказалось, что там в точности этой задачей занимаются уже лет 6–7, причём безуспешно, если не считать алгоритмов для специальных случаев. Я рассказал («проболтался», можно сказать) об этой задаче моему другу Хачияну, с которым мы учились когда-то на Физтехе. Тот никогда раньше двойственностью не занимался, но про сложность знал всё. (Помните специалиста по «теории поиска» Вальдеца из романа Роберта Шекли «Обмен разумов»?) Через месяц задача была решена, точнее, Хачиян (вместе с Микаэлом Фредманом) получил квазиполиномиальный ($n^{\log n}$)-алгоритм [28], который не превзойдён и по сей день; есть ли полиномиальный — неизвестно.

Равновесие Нэша в чистых стратегиях в стохастических и позиционных играх с полной информацией. В обоих случаях моделью игры служит ориентированный граф (орграф) $G = (V, E)$, вершины которого интерпретируются как позиции, а ориентированные рёбра (дуги), выходящие из данной вершины, — как возможные ходы в соответствующей позиции. Позиции разбиваются на $n + 2$ класса $V = V_T \cup V_R \cup V_1 \cup \ldots \cup V_n$. Терминальная позиция не имеет выходящих рёбер; в позиции случая $v \in V_R$ задаётся вероятностное распределение на множестве выходящих ребёр, и ход делается случайно, в соответствии с этим распределением; в позиции $v \in V_i$ ход выбирает игрок i.

Существование РН хорошо известно в двух случаях: (i) когда орграф игры конечен и не имеет ориентированных циклов (орциклов) и (ii) для игр двух лиц с нулевой суммой.

В обоих случаях РН обладает весьма приятными дополнительными свойствами: оно достигается не просто в чистых, но и в *стационарных* (*позиционных*) *равномерно равновесных* стратегиях. Стационарность означает, что ход в любой позиции зависит только от неё, но не от предыстории. Равномерная равновесность означает, что одна и та же ситуация (набор стратегий всех n игроков) является РН при любом выборе начальной позиции.

В случае (i) весьма простое доказательство, основанное на методе обратной индукции, появилось в работах Гарольда Куна [37, 38] и Давида Гейла [29]. Хотя в этих статьях рассматриваются только *терминальные* выигрыши (определённые не на рёбрах орграфа, а лишь на его тупиковых позициях (терминалах), как, например, в шахматной игре) и только «личные» позиции игроков $\{1, \ldots, n\}$, а не позиции случая, но в [29] Гейл упоминает возможность таких обобщений.

Однако ацикличность совершенно необходима. Обратная индукция не проходит при наличии в орграфе игры хотя бы одного орцикла. В то же

время, в шахматной игре такие орциклы имеются (позиция может повторяться), и, тем не менее, Цермело доказал свою теорему, используя условие (ii). Позже, в 1950-е годы, выяснилось, что (ii) позволяет установить наличие седловой точки в чистых стационарных равномерно оптимальных стратегиях в гораздо более широком классе стохастических игр с полной информацией и с *предельным средним эффективным выигрышем*. Это означает, что партия бесконечна и каждый из игроков максимизирует свой выигрыш в расчёте на один ход, иными словами, среднее по Чезаро, т. е. нижний предел среднего выигрыша за первые T ходов при $T \to \infty$.

Этот очень непростой результат был получен усилиями многих авторов: Харди — Литлвуд (1931) [35], Шепли (1953) [65], Жилет (1957) [31], Лигет — Липман (1969) [44]. Доказательство Жилета основано на одной из так называемых тауберовских теорем Харди и Литлвуда о связи сумм Чезаро (предельный средний эффективный выигрыш) и абелевых сумм (дисконтированный выигрыш, не путать с премией) при коэффициенте дисконтирования, стремящемся снизу к $+1$. Жилет применил эту теорему, не проверив выполнения всех её условий; его доказательство было скорректировано позже Лигетом и Липманом.

Отступление. Харди много раз заявлял, что хорошая математика — это чистая математика, исключающая любые приложения. Интересно, что одна из его наиболее абстрактных теорем вовсю используется в теории стохастических игр, набирая многие сотни ссылок.

Опираясь на результат для случая (ii), Тёйсман и Рагхаван [66] доказали существование ситуации РН в чистых стратегиях для любых стохастических игр n лиц с полной информацией. Построенная ими ситуация равномерна, то есть её стратегии не зависят от начальной позиции, однако ход игрока в данной позиции может зависеть от предыстории, то есть стратегии, вообще говоря, не стационарны. Конструкция очень проста. Данной игре n лиц G поставим в соответствие n игр с нулевой суммой: в G_i игрок $i \in I$ противостоит всем остальным игрокам $I \setminus \{i\}$, которые забывают о своих интересах и стремятся минимизировать выигрыш i. Известно, что в каждой такой игре существует седловая точка в чистых стационарных *равномерно оптимальных* стратегиях. (Последнее свойство будет играть главную роль.) Рассмотрим следующую нестационарную ситуацию. Каждый игрок $i \in I$ применяет найденную стратегию x_i до тех пор, пока кто-нибудь из них, скажем j, не отклонится, после чего все остальные переключаются на свои («наказательные») стратегии в игре G_j. Несколько замечаний:

(з0) В принципе, возможно более чем одно отклонение, но в определённой выше ситуации игроки реагируют только на первое и игнорируют все

последующие. Доказывая, что полученная ситуация — РН, мы естественно предполагаем, что только один игрок, j, может отклоняться сколько угодно раз, а все остальные следуют выбранным стратегиям.

(з1) В полученной ситуации стратегии зависят от предыстории, но «не очень сильно». Каждый игрок применяет свою осторожную стратегию, но может однажды переключиться на одну из $n-1$ наказательных.

(з2) Поскольку информация полная, у игроков не будет проблем с определением «нарушителя и момента нарушения».

(з3) Может показаться, что, сделав ход (v, v'') вместо предписанного (v, v'), игрок j получит настолько большую выгоду, что его уже не удастся наказать. Однако это противоречит оптимальности наказательной стратегии коалиции $I \setminus \{j\}$ в игре G_j относительно начальной позиции v, а ведь эта стратегия *равномерно* оптимальна.

(з4) Разумеется, аргументы (з3) верны лишь при некоторых условиях на структуру эффективного выигрыша, которую мы здесь не уточняем. Заметим лишь, что (з3) применимо ко всем типам выигрышей, обычно рассматриваемым в теории стохастических игр.

Если не предполагать (i) или (ii), то РН в чистых стационарных стратегиях может не существовать. Пример на рис. 1 был построен в 1988 г. в [4], см. также [6]. Игра двух лиц задана на полном двудольном 3×3 графе. Игроки 1 и 2 контролируют по три позиции, белые и чёрные соответственно. Отметим, что позиций случая в этой игре вообще нет. Каждое ребро на рис. 1 следует заменить на две противоположно ориентированных дуги,

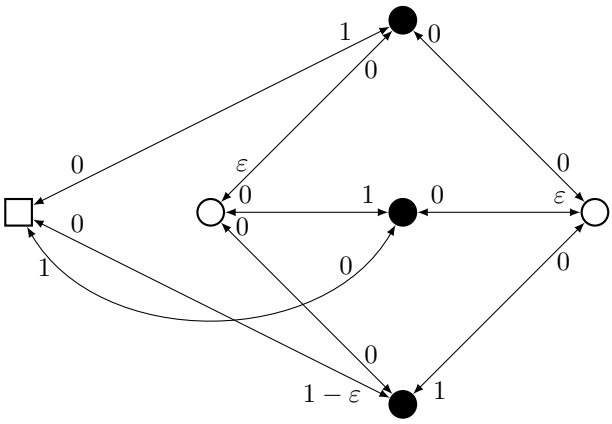


Рис. 1. Игра двух лиц с полной информацией, но с ненулевой суммой без РН в чистых стационарных стратегиях. Локальные выигрыши заданы на рёбрах графа. Белые и чёрные кружки обозначают позиции игроков 1 и 2 соответственно. В начальной позиции ходит игрок 1. Она обозначена белым квадратом

на которых каждый игрок имеет один и тот же локальный выигрыш — число, ближайшее к его вершине. Каждый из игроков имеет $3^3 = 27$ чистых стационарных стратегий. Соответствующая биматричная игра 27×27 не имеет РН в чистых стратегиях.

В [5] показано, что этот пример в некотором смысле минимальный, а именно: любая игра, заданная аналогично на (полном) графе $2 \times k$, имеет РН в чистых стационарных стратегиях.

Пример на рис. 2 был недавно построен в [33]. В этом примере, в отличие от предыдущего, выигрыш терминальный. Это весьма специальный частный случай: все локальные выигрыши на рёбрах для всех игроков равны нулю, а в терминальных позициях (из которых нет ходов), а также для исхода c, соответствующего единственному ориентированному циклу графа, заданы терминальные выигрыши для всех игроков. При этом важна

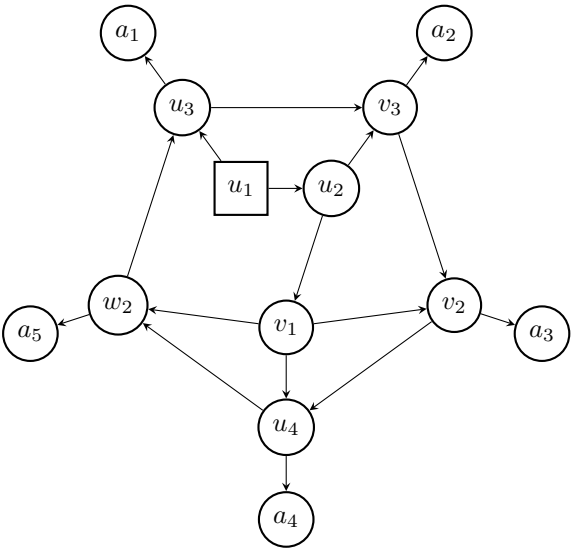


Рис. 2. В этой позиционной игре с полной информацией имеется 4 *игрока,* 6 *исходов:* 5 *терминалов* $\{a_1, \ldots, a_5\}$ *и всего* 1 *цикл* c*. Имеется* 8 *промежуточных позиций. Для каждой нижний индекс обозначает номер игрока, выполняющего ход. Каждый из игроков имеет предпочтение на множестве исходов. Локальных выигрышей нет или, иными словами, все они равны нулю. В* [33] *показано, что игра не имеет РН в чистых стационарных стратегиях всякий раз, когда предпочтения игроков согласуются со следующими частичными порядками*:
$O_1 \colon a_2 \succ a_4 \succ a_3 \succ a_1 \succ a_5$;
$O_2 \colon \min(a_1, c) \succ a_3 \succ \max(a_4, a_5) \succ \min(a_4, a_5) \succ a_2$;
$O_3 \colon \min(a_5, c) \succ a_1 \succ a_2 \succ \max(a_3, a_4)$;
$O_4 \colon \min(a_1, a_2, a_3, a_5) \succ a_4 \succ c$

только их упорядоченность для каждого, а сами вещественные значения роли не играют.

Шахматы, например, — классическая игра с терминальным выигрышем. В ней РН (седловая точка) имеется, поскольку это игра двух лиц с нулевой суммой.

А вот в игре на рис. 2 РН нет. В ней 4 игрока, 5 терминальных позиций и всего один ориентированный цикл. Заметим, что этот цикл c — худший исход для игрока 4, один из двух лучших для 2 и 3, а игроку 1 он вообще безразличен, поскольку у него ходов на цикле нет. Было бы интересно построить такой аналогичный пример, что цикл c — наихудший исход для всех игроков, а также уменьшить число игроков с 4 до 3. Уменьшить его до 2 невозможно; это следует из результатов [2, 3] о разрешимости по Нэшу игровых форм двух лиц; см. стр. 40.

Минимальные матричные игры без седловой точки и биматричные без РН в чистых стратегиях. Легко проверить, что матричная игра 2×2 не имеет седловой точки в чистых стратегиях тогда и только тогда, когда каждый из двух элементов одной из диагоналей строго больше каждого из двух элементов другой.

В 1964 г. Ллойд Шепли [64] показал, что матрица имеет седловую точку тогда и только тогда, когда таковую имеет любая её 2×2-подматрица. Этот результат усилен в [21]: любая матрица без седловой точки размера, превышающего 2×2, имеет строку или столбец, удаление которого не приводит к появлению седловой точки. Иными словами, все (локально) минимальные матрицы без седловой точки имеют размер 2×2.

В [21] этот результат следующим образом обобщается на биматричные игры $a\colon I \times J \to \mathbb{R}$, $b\colon I \times J \to \mathbb{R}$, где a и b — выигрыши игроков 1 и 2 соответственно: (a, b) — локально минимальная биматричная игра без РН (иными словами, (a, b) не имеет РН, но РН появляется после удаления любой строки $i \in I$ или столбца $j \in J$) тогда и только тогда, когда выполнены следующие условия:

(K) квадратность: $|I| = |J| = n$ (но теперь любое значение $n \geqslant 2$ возможно);

(D) существуют такие отображения $\sigma\colon I \to J$ и $\delta\colon J \to I$, что их графики $\mathrm{gr}(\sigma)$ и $\mathrm{gr}(\delta)$ не пересекаются в $I \times J$, иными словами, $(i, \sigma(i)) \neq (\delta(j), j)$ для всех $i \in I$ и $j \in J$;

(A) элемент $a(\delta(j), j)$ — единственный максимум в столбце j и второй по величине (но не обязательно единственный) в строке $\delta(j)$;

(B) элемент $b(i, \sigma(i))$ — единственный максимум в строке i и второй по величине (но не обязательно единственный) в столбце $\sigma(i)$.

Ясно, что может существовать не более одной пары таких отображений (σ, δ).

Достаточность почти очевидна [34]. Пусть условия (K, D, A, B) выполнены. Тогда в игре (a, b) РН нет. В самом деле, для любой ситуации (i, j) имеем: $b(i, \sigma(i)) > b(i, j)$, если $j \neq \sigma(i)$; $a(\delta(j), j) > a(i, j)$, если $i \neq \delta(j)$ и по крайней мере одно из этих двух строгих неравенств выполняется, поскольку иначе $(i, \sigma(i)) = (\delta(j), j) = (i, j)$, вопреки условию (D).

Удалим произвольную строку $i \in I$ и рассмотрим два столбца: $j = \sigma(i)$ и $j' = \delta^{-1}(i)$, строку $i' = \sigma^{-1}\delta^{-1}(i)$ и три ситуации:

$$(i, j) = (i, \sigma(i)), \quad (i, j') = (i, \delta^{-1}(i)), \quad (i', j') = (\sigma^{-1}\delta^{-1}(i), \delta^{-1}(i)).$$

В исходной биматрице $b(i', j')$ — единственный максимум в строке i' и $a(i', j')$ — второй по величине элемент в столбце j', уступающий только $a(i, j')$. Но последний был удалён вместе со строкой i. Значит, в редуцированной биматрице (i', j') — равновесие Нэша. Аналогично если удалить любой столбец $j \in J$, то ситуация $(\sigma^{-1}(j), \delta^{-1}\sigma^{-1}(j))$ станет РН.

Итак, (a, b) — локально минимальная биматрица без РН. Доказать необходимость немного труднее; см. [21].

Отметим, что (a, b) *локально минимальная*, но совсем не обязательно *минимальная* биматрица без РН, поскольку она вполне может содержать подбиматрицу без РН, не пересекающую $gr(\sigma) \cup gr(\delta)$ в $I \times J$. В отличие от локальной минимальности, простой критерий минимальности получить не удаётся.

Абелевская премия была присуждена Нэшу в 2015 г. «за выдающийся и основополагающий вклад в теорию нелинейных уравнений в частных производных и её приложения в геометрическом анализе». В первую очередь, речь идёт о решении 19-й проблемы Гильберта: теореме де Джорджи — Нэша [25, 26, 50, 53, 57]; также получил признание вклад Нэша в вещественную алгебраическую геометрию: теорема Нэша — Тоньоли [58, 67]; теория вложений и погружений: феномен Нэша — Кайпера [43, 52].

19-я проблема Гильберта была решена в нескольких фундаментальных работах. Первый значительный шаг сделал Сергей Натанович Бернштейн в 1903 г. (заняться проблемой предложил ему сам Гильберт), следующий — Иван Георгиевич Петровский в 1937 г. и, наконец, окончательное решение получили независимо Эннио де Джорджи (1956, 1957) [25, 26] и Джон Нэш (1957, 1958) [53, 57].

В 1952 г. Нэш доказал [58], что любое гладкое компактное многообразие диффеоморфно невырожденной компоненте вещественного алгебраического множества; в 1973 г. Альберто Тоньоли [67] усилил эту теорему, избавившись от слова «компонента». Результат получил название «теорема Нэша — Тоньоли»; вошли в обиход также термины функция (отображение), диффеоморфизм, многообразие, крыло Нэша.

Теорема Нэша о регулярных вложениях — это обобщение теоремы об обратной функции с банаховых пространств на пространства Фреше. Всякое m-мерное риманово многообразие (B^m, g) класса C^k, где $3 \leqslant k \leqslant \infty$, допускает изометрическое C^k-вложение в $\mathbb{R}^n$ для достаточно большого n [60]. Нэш дал явную оценку $n \geqslant m(m+1)(3m+11)/2$, которая позднее несколько раз улучшалась; в частности теорема справедлива для $n \geqslant m^2 + 10m + 3$. Нэш также доказал аналогичный результат для аналитических вложений [60].

Для C^1-гладких вложений похожий результат (теорема Нэша — Кайпера [43, 52]) утверждает, что любое гладкое вложение (погружение) n-мерного риманова многообразия в евклидово пространство $\mathbb{R}^q$ при $q > n$ можно аппроксимировать C^1-гладким изометрическим вложением (погружением), где «изометрическое» означает «сохраняющее длины кривых». Нэш получил этот результат при $q > n + 1$.

В 1966 г. Юрген Мозер показал [47], что методы вложения Нэша применимы для решения проблемы периодических орбит в небесной механике (теорема Нэша — Мозера).

Замечание. В 2011 г. Агентство национальной безопасности США рассекретило записки Нэша 1950 г., в которых он предлагает новую шифровальную машину. Записки показывают, что Нэш предвидел многие концепции современной криптографии и, в частности, её связь с теорией сложности вычислений.

«Кампанейщина и волны» в теории игр. Первая и наиболее мощная волна локализована не только во времени (1940–1950), но и в пространстве — Принстон. Такая блестящая компания, наверное, больше нигде и никогда не собиралась. Главным был фон Нейман, затем Шепли, Нэш и их научный руководитель Такер, а также Данциг, Гейл, Кун и Ауман. Можно включить и Какутани, который в 1941 г. в Принстоне доказал свою теорему [36]. Невиданная концентрация! Принстон ведь маленький университет (но, без сомнения, великий). Фон Нейман и Моргенштерн в 1944 г. опубликовали свою знаменитую монографию [69]. Данциг и фон Нейман поняли эквивалентность матричных игр и линейного программирования. Такер, Кун и Гейл обобщили понятие двойственности на случай выпуклого программирования. Шепли ввёл стохастические игры (и, как частный случай, управляемые марковские процессы). В результате теория игр двух лиц с нулевой суммой приобрела вполне законченные очертания.

Тогда же были введены главные понятия теории игр n лиц: некооперативной — РН (1950) и кооперативной — вектор Шепли (1953). Позже появились их многочисленные уточнения и вариации, но они, как правило,

слишком громоздки, чтобы оставить место красивой математике, и всё ещё далеки от реальных приложений, например в экономике. Исключением является, безусловно, понятие ядра, введённое Дональдом Брюсом Джиллисом в 1959 г. [32].

Этот блестящий период закончился около 1960 г., но по инерции теория игр считалась модной ещё лет 15–20. На неё возлагались грандиозные надежды, которые по зрелом размышлении ну никак не могли оправдаться. Я помню, как Михаил Моисеевич Ботвинник в 1975 г. рассказывал, что если его программа «Пионер» заиграет в шахматы хотя бы на уровне мастера, то через год все проблемы экономики будут решены. Психологически понятно, что после чрезмерных надежд наступило чрезмерное же разочарование. Вспомните заявление Алана Хофмана.

Замечание. С последствиями я сталкивался не раз. Например, стало трудно публиковать статьи. Международный журнал теории игр — тоненький и выходил всего 4 раза в год, а остальные журналы значительно уменьшили квоту на игры. Возникла даже самоцензура. В 1980-е годы Леонид Хачиян, Александр Карзанов и я работали над циклическими играми [6]. (Это такой специальный класс «детерминированных стохастических игр». Задача их решения лежит в пересечении классов NP и co-NP, но полиномиального алгоритма до сих пор нет. Выходит, она сложнее задачи линейного программирования, которую Хачиян успешно решил за полиномиальное время в 1979 г. В [6] предложен, в частности, пример циклической игры двух лиц (с ненулевой суммой) без РН.) В 1985 г. я написал первый вариант нашей статьи «Циклические игры», но Хачиян заявил, что игры сейчас не в моде, и всё переписал на языке минимаксных средних циклов в ориентированных графах, добавив их прямо в название статьи и убрав из введения стратегии и «прочие рудименты» теории игр.

Американской прикладной математике и информатике вообще свойственна некоторая «кампанейщина». Время от времени собираются «большие люди» и решают, какие направления на сегодняшний день надо считать основными, а какие, наоборот, притормозить. Затем возглавляемые ими «рядовые исследователи» врываются в указанные приоритетные области, как стадо мустангов в прерии, и, поднимая тучи пыли, начинают свои исследования; причём зачастую «с чистого листа», поскольку кампания рассчитана на определённые (достаточно короткие) сроки и основательно изучать, что там накопилось за десятилетия, времени нет. Как однажды заметил Хачиян: «Потом, когда пыль уляжется, от всей этой деятельности порой и следа не остаётся».

Это произошло и с теорией игр в середине 1990-х годов. Совет тогда возглавили директор Центра дискретной математики и теоретической информатики Фред Робертс, а также Алан Хофман, Христос Пападимитриу и Ева Тардош; Джон Нэш был приглашён в качестве свадебного генерала и даже выступил с коротким докладом. Поскольку всё происходило в Ратгерсе, мы с Хачияном тоже зашли. Было решено, что в связи с бурным развитием электронной торговли и социальных сетей надо теорию игр из изгнания вернуть, а PH назначить основной концепцией, поскольку некооперативная тенденция явно преобладает. Были назначены и сроки мозгового штурма — ближайшие 5–7 лет. Сразу же закипела работа. Число публикаций про PH увеличилось раз в пять. В доброй половине случаев переоткрывались или переформулировались с небольшими отклонениями забытые или просто неизвестные авторам результаты 1970–80-х годов. Появились хлёсткие термины, вроде «цена анархии». (Помнится, в диссертации 1978 г. у меня была «цена стабильности». До сих пор не знаю, это то же самое или нет.)

Действительно большим достижением этой кампании стало понятие *PPAD*-полноты, введённое Пападимитриу [62] ещё в 1994 г., то есть за пару лет до «новой волны». Это понятие напрямую связано с вычислением PH в смешанных стратегиях. Какова алгоритмическая сложность этой задачи? Классы P и NP тут не подходят, поскольку определены только для задач с ответом «да-нет», а неподвижная точка или PH существуют всегда, согласно теоремам Брауэра и Нэша. Вопрос: как их найти? По-видимому, эффективных алгоритмов нет. Пападимитриу формально ввёл класс *PPAD*, аналогичный классу NP, и показал, что обе задачи *PPAD*-полны. Более того, *PPAD*-полнота имеет место уже в случае размерности два для теоремы Брауэра и, соответственно, двух игроков для PH, но доказать это сложнее. Такой результат был получен только в 2006 г. Си Ченем и Сяоте Деном [23].

Вспомним, что дуализация гиперграфов эквивалентна разрешимости по Нэшу игровых форм двух лиц. Для этой задачи в 1996 г. Хачиян предложил квазиполиномиальный алгоритм [28], но полиномиального нет до сих пор. Возможно, задача имеет какой-то особый статус.

Таким образом, PH попадает в самый центр современной теории алгоритмической сложности.

Заключение. В 1971 г. Юрий Борисович Гермейер объяснил мне, что такое PH. С тех пор я написал много статей по теории игр, и половина из них — именно про PH. Я бывал на лекциях Джона Нэша, но никогда с ним лично не беседовал. Конечно, у меня нет права назвать его учителем, но он — человек, построивший мой дом. Жаль, что это понятие не выражается одним словом.

Список литературы

[1] *Гурвич В. А.* К теории многошаговых игр // Ж. вычисл. матем. и матем. физ. 1973. Т. 13, № 6. С. 1485–1500.

[2] *Гурвич В. А.* Равновесие в чистых стратегиях // ДАН СССР. 1988. Т. 303, № 4. С. 538–542.

[3] *Гурвич В. А.* Разрешимость позиционных игр в чистых стратегиях // Ж. вычисл. матем. и матем. физ. 1975. Т. 15, № 2. С. 358–371.

[4] *Гурвич В. А.* Стохастическая игра с полной информацией, не имеющая ситуаций равновесия в чистых стационарных стратегиях // УМН. 1988. Т. 43, вып. 2(260). С. 135–136.

[5] *Гурвич В. А.* Теорема существования ситуаций равновесия в чистых стратегиях для эргодических расширений $(2 \times k)$-биматричных игр // УМН. 1990. Т. 45, вып. 4(274). С. 151–152.

[6] *Гурвич В. А., Карзанов А. В., Хачиян Л. Г.* Циклические игры и поиск минимаксных средних циклов в ориентированных графах // Ж. вычисл. матем. и матем. физ. 1988. Т. 28, № 9. С. 1407–1417.

[7] *Канторович Л. В.* Дальнейшее развитие математических методов и перспективы их применения в планировании и экономике // Применение математики в экономических исследованиях. 1959. С. 310–353.

[8] *Канторович Л. В.* Математико-экономические работы. Избранные труды. Новосибирск: Наука, 2011.

[9] *Канторович Л. В.* Математические методы организации и планирования производства. Л.: ЛГУ. 1939. Перепечатка: в сб. «Применение математики в экономических исследованиях». М.: Социздат, 1959. С. 251–309.

[10] *Канторович Л. В.* О перемещении масс // ДАН СССР. 1942. Т. 37, № 7/8. С. 227–229.

[11] *Канторович Л. В.* Об одном эффективном методе решения некоторых классов экстремальных проблем // ДАН СССР. 1940. Т. 28, № 3. С. 212–215.

[12] *Канторович Л. В.* Показатели работы предприятий нуждаются в пересмотре, 1943, рукопись. Опубл.: Оптимизация (сборник трудов Института математики СО АН СССР). 1991. Вып. 50(67). С. 16–44.

[13] *Канторович Л. В.* Рациональные методы раскроя металла. Произв.-техн. бюл. НКБ СССР. 1942. № 7–8. С. 21–29.

[14] *Канторович Л. В.* Экономический расчёт наилучшего использования ресурсов. М.: Изд-во АН СССР, 1960.

[15] *Канторович Л. В., Залгаллер В. А.* Расчёт рационального раскроя промышленных материалов. Л.: Лениздат, 1951.

[16] *Binmore K.* Playing for real: a text on game theory. Oxford University Press, 2007.

[17] *Borel E.* Eléments de la théorie des probabilités. Paris: Librairie Scientifique, J. Hermann, 1924. P. 204–221. (Англ. пер.: On games that involve chance and the skill of the players // Econometrica. 1953. V. 21. P. 101–115.)

[18] *Borel E.* La théorie du jeu et les équations intégrales à noyau symétrique // Comptes rendus hebdomadaires des séances de l'Académie des sciences. 1921. V. 173. P. 1304–1308. (Англ. пер.: The theory of play and integral equations with skew symmetric kernels // Econometrica. 1953. V. 21. P. 97–100.)

[19] *Borel E.* Sur les systèmes de formes linéaires à déterminant symétrique gauche et la théorie générale du jeu // Comptes rendus hebdomadaires des séances de l'Académie des sciences. 1927. V. 184. P. 52–53. (Англ. пер.: On systems of linear forms of skew symmetric determinant and the general theory of play // Econometrica. 1953. V. 21. P. 116–117.)

[20] *Boros E., Gurvich V.* Why Chess and Backgammon can be solved in pure positional uniformly optimal strategies // Rutgers Research Report 21–2009, Rutgers University.

[21] *Boros E., Gurvich V., Makino K.* Minimal and locally minimal games and game forms // Discrete Math. 2009. V. 309, № 13. P. 4456–4468.

[22] *Boros E., Gurvich V., Makino K., Shao W.* Nash-solvable two-person symmetric cycle game forms // Discrete Applied Math. 2011. V. 159. P. 1461–1487.

[23] *Chen X., Deng X.* On the complexity of 2D discrete fixed point problem // 33rd Int. Colloquium on Automata, Languages and Programming (ICALP, 2006). P. 489–500.

[24] *Cournot A. A.* Recherches sur les principes mathématiques de la théorie des Richesse. Paris: Chez L. Hachette, 1838.

[25] *de Giorgi E.* Sull'analiticita delle estremali degli integrali multipli // Atti della Accademia Nazionale dei Lincei. Rendiconti. Classe di Scienze Fisiche, Matematiche e Naturali. Serie VIII (итал.). 1956. V. 20. P. 438–441.

[26] *de Giorgi E.* Sulla differenziabilita e l'analiticita delle estremali degli integrali multipli regolari // Mem. Accad. Sci. Torino. Classe di Scienze Fisiche, Matematicahe e Naturali. Serie III (итал.). 1957. V. 3. P. 25–43. (Англ. пер.: On the differentiability and the analyticity of extremals of regular multiple integrals // E. de Giorgi. Selected papers. Berlin — New York: Springer-Verlag, 2006. P. 149–166.)

[27] *Edmonds J., Fulkerson D. R.* Bottleneck extrema // J. Combin. Theory. 1970. V. 8. P. 299–306.

[28] *Fredman M. L., Khachiyan L.* On the complexity of dualization of monotone disjunctive normal forms // J. of Algorithms. 1996. V. 21, № 3. P. 618–628.

[29] *Gale D.* A theory of N-person games with perfect information // Proc. Nat. Acad. Sci. 1953. V. 39. P. 496–501.

[30] *Gale D., Shapley L. S.* College admissions and the stability of marriage // Amer. Math. Monthly. 1962. V. 69, № 1. P. 9–15.

[31] *Gillette D.* Stochastic games with zero stop probabilities // Contribution to the theory of games III. Princeton, N. J.: Princeton University Press, 1957. (Ann. Math. Studies; № 39). P. 179–187.

[32] *Gillies D. B.* Solutions to general non-zero-sum games // Contributions to the theory of games IV. Princeton, N. J.: Princeton University Press, 1959. (Ann. Math. Studies; № 40). P. 47–85.

[33] *Gurvich V.* A four-person chess-like game without Nash equilibria in pure stationary strategies // Business Informatics. 2015. V. 1, № 31. P. 68–76. http://arxiv.org/abs/1411.0349.

[34] *Gurvich V., Libkin L.* Absolutely determined matrices // Math. Soc. Sci. 1990. V. 20. P. 1–18.

[35] *Hardy G. H., Littlewood J. E.* Notes on the theory of series (16): two Tauberian theorems // J. London Math. Soc. 1931. V. 6. P. 281–286.

[36] *Kakutani S.* A generalization of Brouwer's fixed point theorem // Duke Math. J. 1941. V. 8, № 3. P. 457–459.

[37] *Kuhn H. W.* Extensive games // Proc. Nat. Acad. Sci. USA. 1950. V. 36. P. 570–576.

[38] *Kuhn H. W.* Extensive games and the problem of information // Contributions to the theory of games. Vol. II. Princeton, N. J.: Princeton University Press, 1953. (Ann. Math. Studies; № 28). P. 193–216.

[39] *Kuhn H. W.* The Hungarian method for the assignment problem // Naval Research Logistics Quarterly. 1955. V. 2. P. 83–97.

[40] *Kuhn H. W.* Variants of the Hungarian method for assignment problems // Naval Research Logistics Quarterly. 1956. V. 3. P. 253–258.

[41] *Kuhn H. W., Harsanyi J. C., Selten R., Weibull J. W., van Damme E., Nash J. F., Jr., Hammerstein P.* The work of John Nash in game theory. A celebration of John F. Nash, Jr. // Duke Math. J. 1995. V. 81:1. P. 1–29.

[42] *Kuhn H. W., Nasar S. (eds.)* The essential John Nash. Princeton University Press, 2001.

[43] *Kuiper N. H.* On C1-isometric imbeddings, I // Nederl. Akad. Wetensch. Proc. Ser. A. 58 = Indag. Math. V. 17. P. 545–556.

[44] *Liggett Th. M., Lippman S. A.* Stochastic games with perfect information and time average payoff // SIAM Review. 1969. V. 11. P. 604–607.

[45] *Milnor J.* A Nobel prize for John Nash // Math. Intell. 1995. V. 17. P. 11–17.

[46] *Milnor J.* John Nash and "A Beautiful Mind" // Notices of the American Mathematical Society. 1998. V. 45. P. 1329–1332.

[47] *Moser J.* A rapidly convergent iteration method and non-linear partial differential equations. I, II // Ann. Scuola Norm. Sup. Pisa. 1966. V. 20, № 3. P. 265–315, 499–535.

[48] *Myerson R. B.* Nash equilibrium and the history of economic theory // J. of Economic Literature. 1999. V. 37. P. 1067–1082.

[49] *Nasar S.* A Beautiful Mind. New York: Simon and Schuster, 1998.

[50] *Nash J. F., Jr.* Analyticity of the solutions of implicit function problem with analytic data // Ann. Math. 1966. V. 84, № 3. P. 345–355. (Рус. пер.: *Нэш Дж.* Аналитичность решений задач о неявной функции с аналитическими исходными данными // УМН. 1971. Т. 26, вып. 4(160). С. 217–226.

[51] *Nash J. F., Jr.* Autobiography. Les Prix Nobel. The Nobel Prizes 1994, ed. T. Fraengsmyr, Stockholm: Nobel Foundation, Almqvist and Wiksell, 1995. P. 275–279.

[52] *Nash J. F., Jr.* C^1-isometric imbeddings // Ann. Math. 1954. V. 60, № 3. P. 383–396. (Рус. пер.: Математика. 1957. Т. 1, № 2. С. 17–28.)

[53] *Nash J. F., Jr.* Continuity of solutions of parabolic and elliptic equations // Amer. J. Math. 1958. V. 80, № 4. P. 931–954.

[54] *Nash J. F., Jr.* Equilibrium points in n-person games // Proc. Nat. Acad. Sci. USA. 1950. V. 36. P. 48–49.

[55] *Nash J. F., Jr.* Non-cooperative games // Ann. Math. 1951. V. 54, № 2. P. 286–295. (Рус. пер.: *Нэш Дж.* Бескоалиционные игры // Матричные игры. М.: Физматгиз, 1961. С. 205–221.)

[56] *Nash J. F., Jr.* Non-cooperative games. Doctoral dissertation, 1950. Princeton University.

[57] *Nash J. F., Jr.* Parabolic equations // Proc. Nat. Acad. Sci. USA. 1957. V. 43, № 8. P. 754–758.

[58] *Nash J. F., Jr.* Real algebraic manifolds // Ann. Math. 1952. V. 56, № 2. P. 405–421. (См. также: Proc. Intern. Congr. Math. AMS, 1952. P. 516–517.)

[59] *Nash J. F., Jr.* The bargaining problem // Econometrica. 1950. V. 18. P. 155–162.

[60] *Nash J. F., Jr.* The imbedding problem for Riemannian manifolds // Ann. Math. 1956. V. 63, № 1. P. 20–63.

[61] *Nash J. F., Jr.* Two-person cooperative games // Econometrica. 1953. V. 21. P. 128–140.

[62] *Papadimitriou Ch.* On the complexity of the parity argument and other inefficient proofs of existence // J. Comp. and System Sciences. 1994. V. 48, № 3. P. 498–532.

[63] *Scarf H.* The core of an N-person game // Econometrica. 1967. V. 35, № 1. P. 50–69.

[64] *Shapley L. S.* Some topics in two-person games // Advances in Game Theory. Princeton, N. J.: Princeton University Press, 1964. (Ann. Math. Studies; № 52). P. 1–28.

[65] *Shapley L. S.* Stochastic games // Proc. Nat. Acad. Sci. USA. 1953. V. 39. P. 1095–1100.

[66] *Thuijsman F., Raghavan T. E. S.* Perfect information stochastic games and related classes // International Journal of Game Theory. 1997. V. 26, № 3. P. 403–408.

[67] *Tognoli A.* Su una congettura di Nash (итал.) // Ann. Scuola Norm. Sup. Pisa. 1973. V. 27, № 1. P. 167–185.

[68] *von Neumann J.* Zur Theorie der Gesellschaftsspiele // Math. Ann. 1928. Bd. 100. S. 295–320. (Рус. пер.: *фон Нейман Дж.* К теории стратегических игр // Матричные игры. М.: Физматгиз, 1961. С. 173–204.)

[69] *von Neumann J., Morgenstern O.* Theory of games and economic behavior. Princeton University Press, 1944. (Рус. пер.: *фон Нейман Дж., Моргенштерн О.* Теория игр и экономическое поведение. М.: Физматгиз, 1970.)

[70] *Zermelo E. F. F.* Ueber eine Anwendung der Mengenlehre auf die Theorie des Schachspiels // Proc. Fifth Congr. Math. Cambridge 1912. V. 2. Cambridge University Press, 1913. P. 501–504. (Рус. пер.: *Цермело Э.* О применении теории множеств к теории шахматной игры // Матричные игры. М.: Физматгиз, 1961. С. 167–172.)

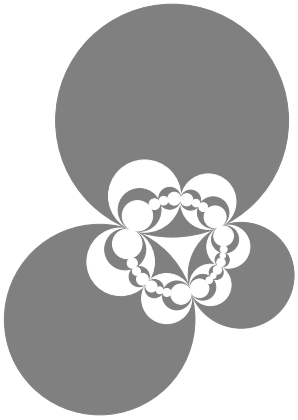
В. А. Гурвич, Ратгерс, Государственный университет Нью-Джерси (США), НИУ ВШЭ

vladimir.gurvich@gmail.com

Геометрия: классика и современность

Тройственная симметрия фрактального калейдоскопа

Р. Р. Пименов



Оглавление

Предисловие 58
§ 0. Три касающиеся окружности в искусстве 60
§ 1. Определение изгиба 62
§ 2. Три сокасающиеся окружности и тройственные координаты 65
§ 3. Точки и калибровка окружности 70
§ 4. Перестановки трёх точек и связанные с ними функции 73
§ 5. Алфавит, слова и композиции симметрий 73
§ 6. Симметрии относительно трёх окружностей в (A–B) координатах. Дробно-линейные преобразования и цепные дроби. Уравнение золотого сечения 81

Математическое просвещение, сер. 3, вып. 20, 2016(57–110)

§ 7. Пары точек и гармоническое отношение 85
§ 8. Подробнее о калибровке окружности. Двойное отношение. Соотношения между изгибами 88
§ 9. Уравнения и предельные точки. Геометрический смысл замены букв в слове 93
§ 10. Поиск закономерностей в знаковых системах. Инверсионная диаграмма 97
§ 11. Подведение итогов и пути развития эстетической геометрии 101
Список литературы 110

Предисловие

Эта статья — развитие и применение идей эстетической геометрии, которыми я стараюсь «заразить» и научное сообщество, и любителей. Она развивает идеи книги [3].

Эстетическая геометрия возникает как способ построения красивых образов на основании симметрий относительно окружности и плодотворна в разных разделах математики, таких как фрактальная геометрия, неевклидовы геометрии, проективная геометрия, основания геометрии. Её основы изложены в электронном учебнике [7] и сопутствующих материалах на сайте. Занимаясь эстетической геометрией, я пришёл к тезису: любое математическое понятие может быть содержательно проиллюстрировано в её рамках, а иллюстрация получается красивой (потому и возникло название «эстетическая» геометрия). Тезис хорошо себя показал на занятиях для школьников по основам теории групп.

В статье изучаются свойства трёх касающихся друг друга окружностей. Конечно, хочется провести ещё четвёртую окружность, касающуюся трёх исходных. Свойства такого «четырёхокружника» разбираются в учебнике [7]. Сейчас мы ограничиваемся изучением трёх касающихся друг друга окружностей, но делаем это разносторонне и порою глубоко. Что здесь можно изучать? Сверх того, что рассказывается о них в учебнике? Что вообще в рамках эстетической геометрии можно построить на основе трёх окружностей?

Можно отражать их симметрично друг относительно друга, получая первые фигуры фрактальной геометрии. Многократные отражения позволяют кодировать слова трёхбуквенного алфавита, моментально зримо представляя их многие важные свойства. Этот метод называется «Инверсионной диаграммой», и он полезен при изучении случайных процессов, а вероятно и шире — в семиотике. Количественный анализ «фрактального калейдоскопа», который при этом получается, требует специальной меры или координат, естественных для геометрии окружности. Для этого в начале статьи излагается понятие «изгиба» или координат, созданных

парой касающихся окружностей. Нам даны три касающиеся окружности, поэтому можно ввести шесть таких систем координат. Правила перехода от одной системы координат к другой просты, элегантны и основаны на *тройственной симметрии* (симметрии третьего порядка). Эта симметрия возникает оттого, что все три исходные окружности абсолютно равноправны, её изучение — простой случай теории групп.

Введя числовую меру для эстетической геометрии, мы можем измерять и решать уравнения. Неожиданно, но первое нетривиальное уравнение, которое удаётся составить в эстетической геометрии, имеет решением золотое сечение, известное из прямолинейной геометрии Евклида. А сама мера изгиба оказывается эквивалентной двойному отношению в проективной геометрии. Проективная геометрия — раздел геометрии прямых, возникший из изучения перспективы во многом под влиянием эстетических устремлений художников и архитекторов. Оказывается, свойства двойного отношения определяются и проявляются в геометрии окружности ярче и естественней, чем в геометрии прямых. Затем уравнения эстетической геометрии записываются через дробно-линейные функции (частное двух линейных функций) и делаются новые приложения к поиску закономерностей в последовательности знаков. Также удаётся численно описать окружности, возникающие во «фрактальном калейдоскопе». Для этого оказываются полезны цепные дроби. В конце статьи намечаются пути развития эстетической геометрии.

Из-за разнообразия появляющихся математических идей и методов возникает мысль, что материал эстетической геометрии годится для прояснения содержательного, неформального единства математики. С конца XIX века единство математики устанавливалось и изучалось на формальном языке: с помощью теории множеств и математической логики. Это привело к бурному развитию математики, но и затруднило взаимопонимание учёных: изобилие и изощрённость формальных, языковых средств сделало современную науку доступной только для узкого круга специалистов. Эстетическая геометрия может изменить это положение, отвечая на полузапретный в мире профессиональных математиков вопрос: «Что же мы на самом деле изучаем в математике?». Я считаю, она — лучший путь к изучению богатейшего мира математики.

Методологически она связана с двумя идеями Лейбница. Его мысль об исчислении, связанном не с внешней системой координат, а с самими изучаемыми фигурами, находит разрешение в исчислении симметрий, постоянно используемом в эстетической геометрии. Подобное делал и Бахман в своих работах по основаниям геометрии на основе симметрий относительно прямых (см. [1]). Вторая связанная с эстетической геометрией мысль Лейбни-

ца — его мысль о монадах. Естественной моделью монады является окружность или сфера, относительно которой отражаются остальные монады-окружности (сферы). Это ясно видно при изучении «фрактального калейдоскопа», который в статье описывается как слова трёх- (или более) буквенного алфавита. Удобное наглядное представление любой последовательности знаков с помощью фрактального калейдоскопа или инверсионной диаграммы указывает на связь геометрии окружности и семиотики.

Сегодня малоизвестны даже основы геометрии окружности. Поэтому — несколько слов о том, как удобней читать эту статью. Все основные понятия подробно изложены в книге [3] и в учебнике [7]. Но я старался, чтобы главные идеи были ясны и без постоянного просмотра учебника. Единственное, что надо знать заранее, — *инверсия, или симметрия относительно окружности*. Если окружность получается из окружности A симметрией (инверсией) относительно окружности B, то мы обозначаем её $B(A)$. Все остальные упоминания теорем эстетической геометрии можно пропустить при первом чтении и вернуться к этим местам позднее, используя уже материал учебника. И, пожалуй, не помешает раньше времени заглянуть в самый конец. Надеюсь, что моя работа не только расскажет об интересных свойствах трёх окружностей и фрактального калейдоскопа, но и послужит введением в высшую математику для любителей. А профессионалы найдут в ней новые связи между известными им теориями.

В известной книге «Ожерелье Индры» [2] инверсии описываются алгебраически, с привлечением вспомогательной техники комплексных чисел и других мощных негеометрических средств. В моей же статье ведётся чисто геометрическая работа. И показано, как на её основе получаются и фундаментальные понятия теории групп, и вычислительные результаты.

Полная версия статьи с цветными иллюстрациями выложена в открытый доступ на сайте [5]. Данная версия статьи адаптирована к условиям публикации в настоящем сборнике. Автор будет рад обсуждению и развитию высказанных в ней идей и результатов. Наука, всякое познание, может твориться в одиночестве, но развивается и цветёт в сообществе.

§ 0. Три касающиеся окружности в искусстве

Поскольку мы занимаемся «эстетической геометрией», начнём изучение трёх сокасающихся окружностей с живописи и орнаментов. Я нашёл не очень много примеров. Лучше всего три касающиеся окружности видны в работах Рериха (см. [9]).

Три и даже четыре касающиеся окружности мы видим на древних кельтских орнаментах, также приведённых в [5].

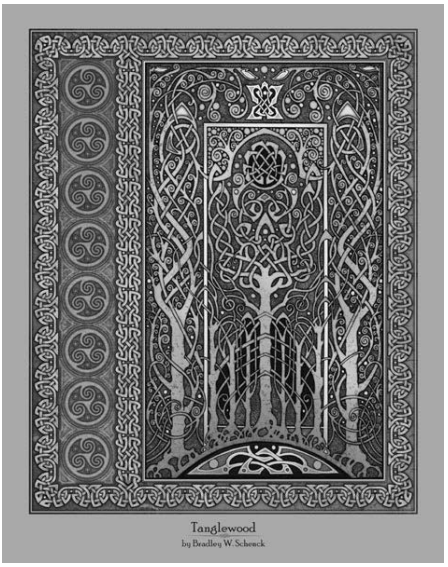


Рис. 1

Рис. 2

Они изображены на левом поле замечательной работы, показанной на рис. 1. Здесь четыре касающиеся окружности, причём они возникают из спиралей. В центре работы сложный узловидный орнамент. Как показано на сайте [6], подобные орнаменты удобно строятся в рамках геометрии окружности, это демонстрирует и программа dodecaLook, имеющаяся на этом же сайте. А на рис. 2 мы видим более простое и подробное изображение четырёх касающихся окружностей-спиралей в кельтском искусстве.

Во всех подобных изображениях три касающиеся окружности явно имеют не только эстетическое значение, но выражают какую-то мистическую или философско-религиозную мысль. Я же буду исследовать их исключительно как математик-геометр.

На рис. 1 и 2 ни одна из трёх маленьких окружностей не лежит внутри другой. В дальнейшем во всей статье предполагается именно такое расположение исходных окружностей A, B, C. Возможен и другой вариант: окружности B и C лежат внутри A. Все наши рассуждения работают и в этом случае, но иногда становятся более громоздкими, поэтому такой вариант расположения окружностей не рассматривается. Читатель может обдумать его самостоятельно (это имеет особый интерес при построении фрактального калейдоскопа).

§ 1. Определение изгиба

Рассмотрим пучок окружностей, определённый окружностями A и B, т. е. совокупность окружностей, касающихся A и B одновременно в точке s. Представим возникновение этого пучка: маленькая, трудноотличимая от точки s окружность X увеличивается в размере — всё время касаясь A в s: вначале она лежит внутри A, в какой-то момент сравнивается с A, касается A с *другой стороны,* продолжает изгибаться и уже сравнивается с B, затем она и B касается с *другой стороны*, затем X уменьшается и снова почти превращается в точку s. На своём пути *из точки в точку* она в какой-то момент становится прямой линией.

В каждый момент окружность X изгибается. Хочется ввести *численную меру* этого изгиба. Чтобы сделать обычную линейку, нам прежде всего надо указать на ней меру: отметки 0 и 1. Чтобы ввести меру изгиба окружности, действуем аналогично: будем считать, что у окружности A — нулевой изгиб, а у окружности B — единичный. Раньше мы считали, что окружность X стартует с какой-то маленькой, неотличимой почти от точки касания окружности. Это наглядно, но не очень удобно. Вводя деления на обычной линейке, нулевую отметку часто помещают в центр линейки. Аналогично мы будем измерять *изгиб* окружности. Окружность A, как мы договорились, имеет нулевой изгиб. Положительный изгиб — это изгиб по направлению к окружности B, самой окружности B мы приписали единичный изгиб. Окружность C, лежащая строго посередине между A и B (т. е. такая окружность, относительно которой A и B симметричны), естественно, получит половинный изгиб. Окружность D, серединная между A и C, имеет изгиб, равный 0,25. И так далее. Окружности с отрицательным изгибом лежат по другую сторону от A. Окружность $A(B)$ имеет изгиб,

равный -1, изгиб $A(C)$ равен $-0{,}5$, изгиб $A(D)$ равен $-0{,}25$ и так далее: если какая-то окружность пучка K имеет изгиб x, то окружность $A(K)$ имеет изгиб $-x$. Заметим: окружность A *разделяет* окружности K и $A(K)$.

Пока все рассмотренные нами окружности имели изгиб от -1 (окружность $A(B)$) до $+1$ (окружность B). Где лежат окружности большего изгиба? Они лежат по другую сторону от B. Например, окружность $B(A)$ имеет изгиб 2. Если мы продолжим сгибать окружность в том же направлении, её изгиб будет возрастать. Если же мы будем сгибать окружность A в *другом* направлении, то получим сначала окружность $A(B)$ с изгибом -1. При продолжении работы изгиб будет всё большим по абсолютной величине, но отрицательным по знаку, а окружность будет всё меньше отличаться от точки s, как это было вначале. Но мы приходим к этой точке *с другой стороны*.

Мы можем представлять кузнеца за работой: перед ним прут, зажатый в точке s. В начальный момент пусть прут совпадает с окружностью A, кузнецу вовсе не надо изгибать прут, чтобы получить A, поэтому изгиб A и назван нулевым; усилие кузнеца, нужное, чтобы изогнуть прут до окружности B, мы считаем единичным; работая дальше и прикладывая всё больше сил, кузнец сгибает прут в точку; сгибая прут в другую сторону, он в конце концов получает то же самое.

На рис. 3 мы видим несколько окружностей из пучка касающихся окружностей (A, B). Окружность F имеет отрицательный изгиб, окружности C и D имеют положительный изгиб, меньший чем 1, окружность E имеет изгиб, больший чем 1.

После инверсии относительно произвольной окружности S с центром в q все окружности пучка перейдут в семейство параллельных прямых (рис. 4). Причём благодаря разобранному в учебнике [3] принципу «симметричное

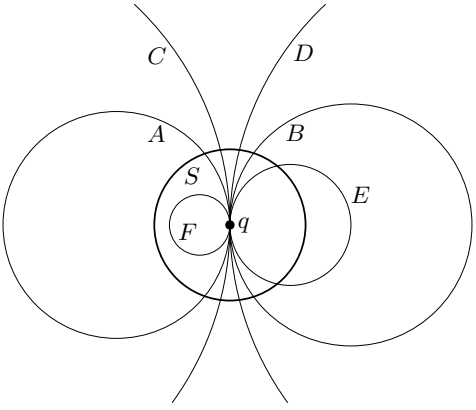


Рис. 3

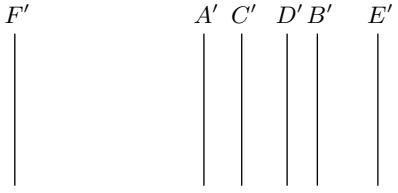


Рис. 4

симметрично симметричному» есть простое соответствие расстояний между получающимися прямыми и $(A–B)$ изгибом окружностей, а именно: $S(F) = F'$, $S(A) = A'$, $S(C) = C'$, $S(D) = D'$, $S(B) = B'$, $S(E) = E'$.

Например, $(A–B)$ изгиб окружности C есть отношение расстояния между прямыми A' и C' к расстоянию между A' и B'. Аналогично можно определить изгиб других окружностей пучка. При этом прямые, лежащие слева от A', мы считаем лежащими в отрицательной области (как и на линейке точки слева от нуля мы считаем отрицательными). Разумеется, так происходит потому, что прямая B' лежит справа от A'. Заметим ещё, что изгиб окружностей не меняется при симметрии относительно произвольной окружности W (не обязательно с центром в q). Это опять-таки легко следует из принципа «симметричное симметрично симметричному» (или из рассмотрения автоморфизма, определённого симметрией относительно W).

Назовём введённый изгиб «$(A–B)$ координатой» окружности пучка. И введём обозначение: пусть $(A–B)(X)$, где A, B — касающиеся окружности, а X — окружность из созданного ими пучка, означает изгиб окружности X относительно пары окружностей A и B. Мы можем обобщить наше обозначение на случай, когда рассматривается не окружность, а произвольная точка плоскости. Ведь через точку x можно провести единственную окружность X, касающуюся A и B в точке s. Именно $(A–B)$ координату этой окружности X мы назовём и $(A–B)$ координатой *точки* x. Особый случай, когда x совпадает с s, мы можем исключить из рассмотрения, а можем считать в этом случае $(A–B)(x)$ равным бесконечности. Отметим, что каждая окружность пучка *однозначно* характеризуется своим изгибом, а точки плоскости — нет. Возьмём произвольную точку x и проведём через неё окружность X из пучка, созданного A и B. Все точки, лежащие на этой окружности, имеют тот же изгиб, что и точка x.

Пусть мы знаем $(A–B)$ координату окружности (или точки) X. Как найти $(B–A)$ координату X? Пучок тот же самый, изменились лишь условия подсчёта изгиба: теперь окружность B имеет нулевой изгиб (кузнец начинает гнуть окружность B, а не A), а окружность A — единичный изгиб. Обозначим исходный изгиб z, а искомый $h(z)$. Функция h должна обладать *двойственной симметрией* (симметрией второго порядка), иначе говоря, быть инволютивной: $h(h(z)) \equiv z$. Подсчитаем эту функцию для нескольких простых случаев: $h(0) = 1$, $h(1) = 0$, $h(0{,}5) = 0{,}5$. В самом деле, нулевой изгиб был у окружности A, но после переобозначения ей присвоен единичный изгиб; единичный изгиб был у окружности B, после переобозначения у неё нулевой изгиб. Половинный изгиб имела серединная окружность C, она и останется серединной, если A и B поменять местами. Возникает гипотеза,

что $h(z) = 1 - z$. Это можно доказать, рассмотрев инверсию с центром в s и применив тезис «симметричное симметрично симметричному». Запишем этот вывод так: $(A–B)(X) = 1 - (B–A)(X)$ или $(A–B)(X) + (B–A)(X) = 1$.

Мы столкнулись с двойственной симметрией пучка окружностей. Теперь мы перейдём к тройственной симметрии.

§ 2. Три сокасающиеся окружности и тройственные координаты

Пусть три окружности A, B, C касаются друг друга. Три точки их касания обозначим s, p, q. При этом A и B касаются в точке s, B и C — в точке p, A и C — в точке q. Каждая пара окружностей определяет описанную выше систему координат для измерения изгиба. Если окружность X лежит в пучке (A, B), то мы можем измерить её $(A–B)$-изгиб, но тогда она не лежит ни в пучке (B, C), ни в пучке (A, C) и мы не можем определить её $(B–C)$ или $(A–C)$ координаты. Но у *точки* плоскости x можно найти три координаты: $(A–B)(x)$, $(B–C)(x)$ и $(C–A)(x)$, проводя через x окружности, лежащие в нужных пучках. Таким образом, каждой точке плоскости можно сопоставить три числа: её $(A–B)$, $(B–C)$ и $(C–A)$ координаты. Можно ли, зная одно из этих чисел, определить два других? Нет.

На рис. 5 изображены точка x и три проходящих через неё окружности. Одна лежит в пучке (A, B), другая в пучке (B, C), третья в пучке (A, C). Обратите внимание, что третья окружность обязательно проходит через точку пересечения двух других.

Если мы сдвинем точку x по большой правой окружности, то её $(A–C)$ координата не изменится, а $(A–B)$ и $(B–C)$ координаты изменятся. Поэтому

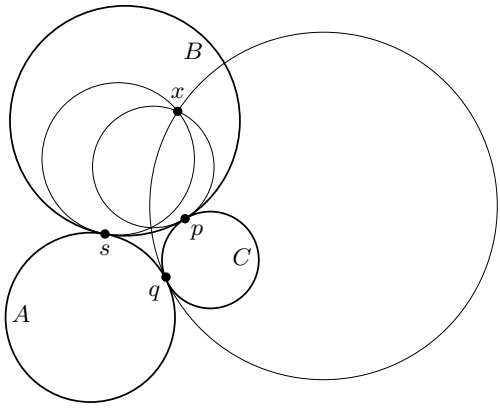


Рис. 5

не может быть способа определить по одной координате точки x две другие её координаты.

Но мы будем изучать $(A–B)$ координаты не всех точек плоскости, а только некоторых. Проведём окружность I через точки s, p, q. Она перпендикулярна трём исходным окружностям A, B, C. Именно $(A–B)$ координаты точек, лежащих на этой окружности, мы и будем изучать. Через произвольную точку x, лежащую на I, мы проводим окружность X, лежащую в пучке (A, B), и изгиб X в этом пучке и будет $(A–B)$ координатой точки x. Легко видеть, что *различные* точки x и y, лежащие на I, обязательно имеют разные $(A–B)$ координаты — поэтому точка на I однозначно задаётся своими координатами. Мы уже говорили о кузнеце, стягивающем и разгибающем прут. В каждый момент этот «прут» (окружность пучка (A, B) пересекает I в двух точках: одна точка — s, точка касания A и B, она общая для всех «прутьев», а вторая точка пробегает всю окружность I, двигаясь через A и B.

Чтобы разобраться с этой и некоторыми подобными конструкциями, удобно воспользоваться стереографической проекцией, точнее её простейшим, одномерным случаем. Стереографической проекцией сферы называется проекция сферы на плоскость (на которой лежит сфера) из высшей точки сферы. В нашем случае не сфера проектируется на плоскость, а окружность на прямую, как изображено на рис. 6: окружность I проектируется на касающуюся её горизонтальную прямую из точки s. Точка s — высшая точка окружности, диаметрально противоположная точке касания прямой и окружности.

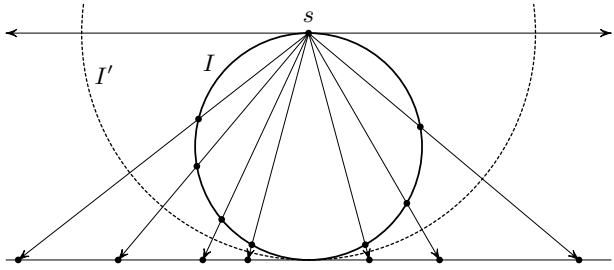


Рис. 6

Точке окружности ставится в соответствие точка на прямой. Но для самого центра проектирования s такой точки нет. Мы можем считать, что точке s соответствует бесконечно удалённая точка на прямой. Заметим, что такое же соответствие точек задаёт и инверсия относительно окружности I' с центром в s, касающейся прямой проектирования.

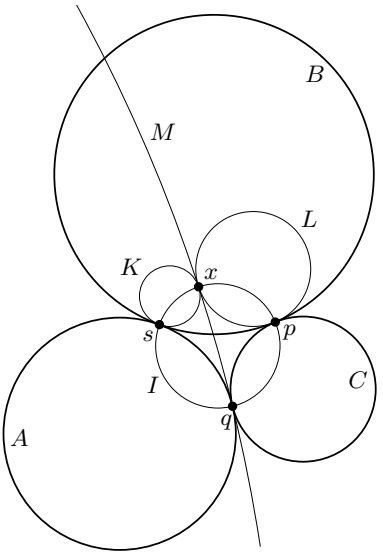


Рис. 7

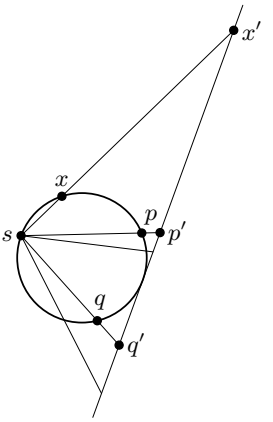


Рис. 8

Мы построили окружность I по трём сокасающимся окружностям A, B, C и выяснили, как определить $(A–B)$ координату произвольной точки x на I. Но точка x на I имеет не только $(A–B)$ координату, но и $(A–C)$ и $(C–B)$ координаты. Как связаны эти координаты между собой?

Через точку x на I проведены (см. рис. 7) три окружности K, L, M, лежащие в соответствующих пучках (между собой эти окружности обязательно касаются, что следует, например из теоремы о трёх окружностях). Интересующие нас изгибы — это $(A–B)(K)$, $(B–C)(L)$ и $(A–C)(M)$.

Как уже говорилось, значение каждого изгиба однозначно задаёт точку x на I. Величину изгиба легко представить с помощью стереографической проекции. Сделаем такую проекцию из точки s (см. рис. 8).

Аналогично связи между изибом окружностей и расстоянием между параллельными прямыми есть связь между $(A–B)$ координатой точки x и расстояниями между точками на прямой проектирования. Именно, $(A–B)(x) = |q', x'| : |q', p'|$. Это можно уяснить из принципа «симметричное симметрично симметричному». В данном случае речь идёт о связи (изоморфизме) между симметриями окружностей в пучке и симметрией точек на прямой проектирования друг относительно друга. Спроектировав из точек p и q на соответствующие прямые, мы получим два других числа для изгибов, измеренных в других системах координат.

Введём обозначения:

$$(A–B)(x) = k, \quad (A–C)(x) = f(k), \quad (B–A)(x) = h(k).$$

Для измерения изгиба мало указать пучок (две касающиеся окружности), нужно ещё указать порядок, направление изгибания. Поэтому каждая пара окружностей определяет две возможные системы координат («прямую» и «обратную»), всего получается 6 возможных систем координат. И, зная хоть одно из этих шести чисел, можно вычислить оставшиеся пять возможных координат. Как это сделать, какая связь между координатами — мы и будем сейчас выяснять.

Об одной связи мы уже говорили:

$$(A–B)(x) = 1 - (B–A)(x) \quad \text{или} \quad (A–B)(x) + (B - A)(x) = 1.$$

Эта формула была указана ранее для случая, когда x — окружность, она очевидно верна и если x — точка. Поэтому $h(k) = 1 - k$. Но пока мы будем писать всюду $h(k)$, а не $1 - k$, потому что так ясней видны закономерности, связывающие координаты. Мы заметили, что $h(h(k)) = k$. Аналогично $f(f(k)) = k$: функция перехода от $(A–B)$ координаты к $(A–C)$ координате должна быть такой же, как функция перехода от $(A–C)$ координаты к $(A–B)$ координате. Через функции f и h можно выразить координаты точки x в любой системе координат, которую мы составим из трёх исходных окружностей.

Пусть мы хотим выразить $(B–C)(x)$ через $(A–B)(x) = k$. Применим сначала преобразование h, меняющее координаты местами. Мы положили $h(k) = (B–A)(x)$. Выразить $(B–C)(x)$ через $(B–A)(x)$ просто: преобразование f как раз и заменяет *вторую* координату, потому $(B–C)(x) = f(h(k))(x)$, что и требовалось. А как заменить *первую* координату? Поступим так: перевернём координаты $(A–B)$, заменим *вторую* координату и снова перевернём. «Переворачивание» координат обеспечивает функция h, *вторую* координату заменяет f, поэтому $(C–B)(x) = h(f(h(k)))$. Аналогично можно поступать и в других случаях. Сведём результаты в таблицу.

Система координат	$(A–B)(x)$	$(B–A)(x)$	$(A–C)(x)$	$(C–A)(x)$	$(B–C)(x)$	$(C–B)(x)$
Значение	k	$h(k)$	$f(k)$	$h(f(k))$	$h(f(h(f(k))))$	$f(h(f(k)))$

Мы видим, что взаимосвязь между разными координатами выражается с помощью композиций всего двух функций: f и h. Функции f и h инволютивны, то есть дважды применённые дают тождественное преобразование: $f(f(k)) = h(h(k)) = k$. А что происходит при их композиции? Если мы применим функцию f, то вторая окружность будет заменена оставшейся окружностью. А после применения h первая координата станет второй, а вторая — первой. В общем итоге на первом месте будет стоять оставшаяся окруж-

ность, а на втором — стоявшая на первом: $(C–A)(x) = h(f(A–B)(x))$. Композиция, применённая трижды, ничего не меняет. Положим $w(k) = h(f(k))$ (или, используя специальный знак для композиции, $h * f = w$). При этом $w(w(w(k))) = k$ (или $w * w * w = e$, где e — тождественная функция). Так мы пришли к тройственной симметрии. Выпишем соотношения подробнее:

$$w((A–B)(x)) = (C–A)(x),$$
$$w((C–A)(x)) = (B–C)(x),$$
$$w((B–C)(x))(A–B)(x).$$

Напомним, что x — точка, а k — число, координата точки x, и все введённые функции w, f, h действуют на числах. Каждый раз переход происходит по следующему правилу: на место первой окружности ставится та, которой нет среди двух используемых, а та, которая была первой, ставится на второе место. Если мы представим A, B, C точками на окружности, то w можно мыслить просто сдвигом.

Но чему равна f — функция, обеспечивающая переход от $(A–B)$ координаты к $(A–C)$ координате? Перебирая разные простые функции, а в качестве первых гипотез всегда стоит пробовать простые варианты, мы натыкаемся на функцию $y = 1/z$. Эта функция инволютивна. Рассматривая точки на I, заметим, что точка с $(A–B)$ координатой 0 (точка q) имеет $(A–C)$ координату, равную бесконечности. Точка с $(A–B)$ координатой 1 (точка p) имеет $(A–C)$ координату 1. Наконец, точка с $(A–B)$ координатой бесконечность (точка s) имеет $(A–C)$ координату 0. Но именно таковы значения функции $y = 1/z$. Далее, как мы показали, композиция $w(z) = (h * f)(z) = h(f(z))$ обладает тройственной симметрией: $w(w(w(z))) = z$ для всех z. При этом $h(z) = 1 - z$. Проверим, удовлетворяет ли функция $y = 1/z$ этому тождеству. Подставим $h(z) = 1 - z$ и получим: $w(z) = 1 - 1/z$. В этом случае в самом деле имеется тождество $w(w(w(z))) = z$. Итак, функция $y = 1/z$ — очень хороший кандидат на роль f. Строгое доказательство того, что $f(z) = 1/z$, можно провести либо с помощью проектирования, о котором говорилось недавно (но это требует некоторой вычислительной работы), либо с помощью связи между заменой координат и симметрией относительно биссектрисы, о которой будет сказано в § 9. Заметим, что в поисках функции f мы наткнулись на важное и интересное тройственное тождество для композиции функций $1 - z$ и $1/z$, функции $w(z) = 1 - 1/z$, которое почему-то не часто упоминают при обучении математике. Разумеется, аналогичное тождество есть и для функции, обратной к w: $y(z) = 1/(1 - z)$. Изучение этих и связанных тождеств — прекрасное введение в теорию групп. Теперь мы составим таблицу перехода от одной системы координат к другой, используя найденный вид функции f.

Система координат	$(A–B)(x)$	$(B–A)(x)$	$(A–C)(x)$
Значение	k	$h(k) = 1 - k$	$f(k) = \frac{1}{k}$
Система координат	$(C–A)(x)$	$(B–C)(x)$	$(C–B)(x)$
Значение	$h(f(k)) = 1 - \frac{1}{k}$	$h(f(h(f(k)))) = \frac{1}{1-k}$	$f(h(f(k))) = \frac{k}{k-1}$

Полезно проверить эти несложные вычисления самостоятельно. Очевидно, $h(f(h(f(h(f(k))))))=k$ (это просто перезапись того, что $w(w(w(k)))=k$), из чего следует ряд тождеств, например: $h(f(h(k))) = f(h(f(k)))$. А вот факт, что функция $v(k) = h(f(h(k))) = 1 - 1/(1 - k)$ тоже инволютивна, т. е. $v(v(k)) = k$, можно вывести уже из того, что f и h инволютивны.

Наши функции f и h и все функции, которые можно получить их композициями (включая тождественную функцию), в совокупности образуют группу. Отметим — идея рассмотрения одного и того же объекта в разных координатах играет ключевую роль во многих разделах математики и физики. Без неё немыслимо векторное исчисление, а теория относительности даже получила своё название из-за неё: она исходит из того, что один и тот же равномерно двигающийся предмет рассматривается разными, также равномерно двигающимися, наблюдателями. Правила перехода от одного наблюдателя к другому, как и здесь, образуют группу, а свойства этой группы и лежат в основе специальной теории относительности. Рассматриваемая смена координат, образованных двумя окружностями, — простейший, но и значительный случай применения этого метода. Кстати, между геометрией окружности и геометризацией теории относительности (метод Минковского) есть прямая связь, но эта тема выходит за рамки нашей статьи.

§ 3. Точки и калибровка окружности

Три касающиеся друг друга окружности A, B, C можно задать, указав точки их касания: s, p, q. Действительно, проведём через эти точки окружность I. Мы это уже сделали: именно на I лежат точки x, $(A–B)$ координаты которых мы изучали. Возьмём из трёх точек s, p, q какие-нибудь две и проведём через них окружность, перпендикулярную I. Это можно сделать одним и только одним способом. Мы получим одну из трёх данных окружностей. Так же строятся и две другие окружности.

Выберем упорядоченную пару окружностей, например (A, B). Мы можем обозначить эту пару окружностей *упорядоченной* тройкой точек: нач-

нём с точки q, в которой A касается C, на второе место поставим точку s, в которой A касается B, а на третье — точку p, в которой B касается C. Таким образом, паре окружностей (A, B) соответствует последовательность точек (q, s, p). Мы можем по упорядоченной тройке точек однозначно восстановить упорядоченную пару окружностей (проходящих через эти точки). Именно: проведём окружность, перпендикулярную I, через две первые точки (q и s), она будет первой окружностью искомой пары. Вторую окружность мы проведём также ортогонально I через вторую и третью точки (s, p) данной тройки.

Теперь мы можем говорить не только про (A, B) координаты произвольной точки x, но и про (q, s, p) координаты точки x. Мы можем ввести обозначение $(q, s, p)(x)$, где q, s, p, x — произвольные точки, лежащие на одной окружности. Тогда $(q, s, p)(x)$ есть изгиб окружностей, касающихся друг друга в точке s (второй точки в тройке), точнее говоря, $(q, s, p)(x) = (A, B)(X)$, где A — окружность, проходящая через (q, s), B — проходящая через (s, p), X — проходящая через (s, x). Все названные окружности, естественно, должны быть перпендикулярны I. Важно, что переход от окружностей к точкам оказался возможен только потому, что любая пара точек на окружности I однозначно определяет окружность, проходящую через эту пару и перпендикулярную I.

Определив координаты с помощью трёх упорядоченных точек, вернёмся к вопросу о «замене координат». Ранее для этого у нас были функции f и h, где h меняет местами координаты, f заменяет вторую окружность на оставшуюся. Эти две функции показывали, как меняются значения координат фиксированной точки x. Ранее координаты определялись упорядоченной парой окружностей, теперь они выражаются упорядоченной тройкой точек. Укажем связь в таблице: к уже используемой таблице добавим строку, выражающую координаты в точечном виде.

A и B касаются в s, B и C в p, A и C в q						
Система координат, *окружности*	$(A$–$B)(x)$	$(B$–$A)(x)$	$(A$–$C)(x)$	$(C$–$A)(x)$	$(B$–$C)(x)$	$(C$–$B)(x)$
Система координат, *точки*	$(q, s, p)(x)$	$(p, s, q)(x)$	$(s, q, p)(x)$	$(p, q, s)(x)$	$(s, p, q)(x)$	$(q, p, s)(x)$
Значение координаты	k	$h(k)$	$f(k)$	$h(f(k))$	$h(f(h(f(k))))$	$f(h(f(k)))$

В точечном виде удобно объяснить смысл функций h и f: функция h определяет, как меняется координата фиксированной точки x, если мы меняем местами первую и последнюю точки (среди трёх выбранных q, s, p); функция f определяет, как меняется координата точки x, если мы меняем местами первую и вторую точку. Какой геометрический и алгебраический смысл в этих заменах и точках q, s, p? Посмотрим на первый столбец таблицы.

Первая точка, q, — точка, где окружность A пересекает I и касается C. Окружность A в пучке (A, B) имеет нулевую координату, также и точка q имеет координату, равную нулю. Вторая точка, s, — точка, где A и B касаются. Как мы говорили, эта точка имеет бесконечную $(A$–$B)$ координату, она и есть *бесконечно удалённая* точка. Наконец, третья точка, p, — точка пересечения B и I. Окружность B в пучке (A, B) имеет единичную координату, потому p также имеет единичную координату. Это можно выразить так: $(q, s, p)(q) = 0$, $(q, s, p)(s) = \infty$, $(q, s, p)(p) = 1$. Процесс установления координат на окружности I (калибровки окружности) оказывается прост: мы можем любую точку на окружности *назвать* нулём (началом). Затем любую другую точку *назвать* бесконечно удалённой, и, наконец, выбрать третью точку и *назвать* её единицей (мерой). От калибровки прямой (или разметки линейки) это отличается тем, что на прямой надо указать *две* точки (нуль и единицу), а на окружности — *три* (нуль, единицу и бесконечность). Можно считать, что прямая — это окружность, на которой одна точка (бесконечность) уже выделена.

Если мы фиксируем точку x и как-то меняем на окружности положение точек, названных нулём, единицей и бесконечностью (точек p, q, s), то будет меняться и значение координаты точки x (так как меняется сама система координат). Если на линейке мы поменяем местами точки 0 и 1, то как изменится координата k фиксированной точки x? Она станет равной $1 - k$. У нас такой сменой заведует функция h. Калибровка прямой достигается указанием всего *двух* точек, поэтому есть всего одна перестановка: они меняются местами. Калибровка окружности достигается указанием *трёх* точек, а три точки можно переставлять по-разному, потому и появляется функция f и композиции функций h и f.

Обратим теперь внимание, как переставляются (p, q, s) координаты под действием w. Это закрепит наше знакомство с тройственной симметрией и выражением её в символическом виде:

$$w((p, q, s)(x)) = (s, p, q)(x),$$
$$w((s, p, q)(x)) = (q, s, p)(x),$$
$$w((q, s, p)(x)) = (p, q, s)(x).$$

Функция w определяет изменение координаты точки при такой «перезаписи» точек: третья точка становится первой, первая — второй, вторая — третьей.

§ 4. Перестановки трёх точек и связанные с ними функции

Итак, у нас есть три точки, которые мы *называем* Нуль, Единица и Бесконечность. А введённые функции f и h (и их композиции) управляют тем, как меняются координаты фиксированной точки x в результате того, что мы *переименовываем* три выбранные точки. То, что мы раньше называли *переименованием,* теперь будем называть *перестановкой* элементов A, B, C. А как записывать перестановку? Пусть, например, первая буква тройки указывает, куда переходит A, вторая — куда переходит B, третья — куда переходит C. Есть и другой удобный способ записи. Пусть мы хотим выразить какую-то перестановку w. Возьмём букву A, следующей запишем букву $w(A)$, в которую A переходит, затем букву $w(w(A))$, в которую переходит $w(A)$, и так далее. Скажем, последовательность (A, B, C) теперь определяет перестановку: A переходит в B, B переходит в C, C переходит в A. Последовательность (A, B) выражает перестановку A и B.

Заметим, что в теории представлений обычно изучают представление групп линейными преобразованиями евклидова пространства. Здесь мы получили представление группы перестановок трёх элементов преобразованиями окружности.

§ 5. Алфавит, слова и композиции симметрий

Подготовив базу, то есть систему координат или меру изгиба, подходящую для количественного анализа фрактального калейдоскопа, начнём строить калейдоскоп и анализировать его качественно. Мы рассматриваем три сокасательные окружности A, B, C. Что будет происходить при композиции симметрий относительно них? Возьмём произвольную точку x, лежащую вне исходных окружностей A, B, C. Будем отражать её относительно A, B, C в каком-то порядке, получим последовательность точек, например: x, $A(x)$, $B(A(x))$, $C(B(A(x)))$, $B(C(B(A(x))))$, … С каждым отражением точка x приближается к окружности I. Подобную картину мы также видим при отражениях относительно прямых (точнее плоскостей), а не окружностей (или сфер).

Пусть имеется последовательность симметрий относительно A_1, A_2, $A_3, \ldots, A_k$, где все окружности A_i — одна из трёх исходных окружностей A, B или C. Заметим, что в этой последовательности мы считаем все

соседние окружности различными. Ведь если рядом стоят две одинаковые окружности, то композиция симметрий относительно них — тождественное движение и их обе можно просто исключить из последовательности. Докажем, что результат композиции симметрий относительно этих окружностей обязательно лежит внутри *последней* окружности, окружности A_k. В самом деле, исходная точка x (напомним, она лежит вне окружностей A, B, C) после симметрии относительно одной из окружностей обязательно лежит внутри неё. Но тогда x будет *вне* двух оставшихся. Поэтому после симметрии относительно какой-то из этих двух точка окажется *внутри* неё (и вне двух оставшихся). И так далее — точка оказывается всё время внутри той окружности, относительно которой проведена последняя симметрия. Что и требовалось.

Пусть теперь имеется вторая последовательность $B_1, B_2, B_3, \ldots, B_s$, где все окружности B_i — также A, B или C, и мы знаем, что

$$A_k\big(A_{k-1}(\ldots(A_2(A_1(x)))\ldots)\big) = B_s\big(B_{s-1}(\ldots(B_2(B_1(x)))\ldots)\big).$$

Что можно сказать про A_i и B_i? При каких условиях во втором случае точка x попадёт туда же, куда и в первом? Как было доказано,

$$A_k\big(A_{k-1}(\ldots(A_2(A_1(x)))\ldots)\big)$$

обязательно лежит внутри A_k, а

$$B_s\big(B_{s-1}(\ldots(B_2(B_1(x)))\ldots)\big)$$

— внутри B_s. Поэтому равенство возможно, только если $A_k = B_s$. Так как при симметрии разные точки переходят в разные, из $A_k = B_s$ следует

$$A_{k-1}(\ldots(A_2(A_1(x)))\ldots) = B_{s-1}(\ldots(B_2(B_1(x)))\ldots).$$

Но из этого равенства следует, что $A_{k-1} = B_{s-1}$. И так далее. Мы получаем, что равенство возможно, только если справа и слева стоят одинаковые окружности, т. е. $k = s$ и $A_i = B_i$.

Рассуждение нетрудно обобщить. Пусть у нас есть не три сокасающиеся окружности, а несколько таких непересекающихся окружностей, что любые две из них лежат по одну сторону от любой другой. Можно представлять их небольшими окружностями-монетами, такими, что ни одна не лежит внутри другой. Как и раньше, возьмём произвольную точку x (лежащую вне всех окружностей), отразим её относительно какой-то одной из исходных окружностей, результат — относительно другой, и так далее. Точка x при этом путешествует по внутренности исходных окружностей, каждый

раз находясь внутри той окружности, относительно которой была последняя симметрия. Как и в рассмотренном ранее случае трёх сокасающихся окружностей A, B, C, если в ходе какой-то последовательности симметрий точка x попала туда же, что и в ходе другой последовательности симметрий, то обе эти последовательности симметрий совпадают. Доказательство аналогично.

Как удобно записывать эти композиции? Назовём исходные окружности (например, A, B, C, D, E) *алфавитом*. Тогда каждое *слово*, составленное из этих букв, можно *прочитать* как последовательность симметрий. Например, $ABCDAE$ читается как композиция симметрий: сначала относительно E, потом относительно A, потом относительно D и так далее по записи. Обратите внимание — мы читаем слово *справа налево*. Если в каком-то слове рядом стоят две одинаковые буквы — их можно «сократить», то есть удалить обе эти буквы из слова. Очень важно, что два разных слова задают *различные* преобразования: это следует из того, что, как было показано, даже на одну точку x, лежащую вне всех окружностей, включённых в алфавит, разные слова действуют по-разному. Заметим, что если бы среди окружностей алфавита были пересекающиеся, то разные композиции относительно пересекающихся окружностей *могли бы* дать один и тот же результат. Например, если P и Q перпендикулярны, то $P * Q = Q * P$ и слово PQ означает то же преобразование, что и QP.

Как выглядит слово, обратное данному (то есть задающее обратное преобразование)? Нужно просто записать данное слово «задом наперёд», перевернуть его. При этом выделится совокупность слов, которые одинаково читаются слева направо и справа налево: *симметричные* слова, например A, ABA, $ABCBABCBA$ и так далее. Какие преобразования плоскости задают такие слова? Очевидно, это *инволютивные* преобразования, т. е. если симметричное слово задаёт преобразование f, то $f * f = e$ (e — тождественное преобразование). Все симметрии инволютивны. Появляется догадка, что симметричные слова означают какие-то симметрии на плоскости. Какие именно? Разберёмся, как выглядит симметричное слово. Предположим, что в нём чётное число букв, $2k$. Тогда буквы с номерами k и $k + 1$ должны быть одинаковы. Но две соседние одинаковые буквы можно удалить из слова, после их удаления рядом снова окажутся одинаковые буквы. И так далее — все буквы в симметричном слове чётной длины будут сокращаться. Пустое слово можно считать тождественным преобразованием e. Перейдём к словам нечётной длины $2k + 1$. У них имеется буква, стоящая в середине слова на $(k+1)$-м месте. Буквы, стоящие на равном расстоянии от неё, должны совпадать. Это даёт лёгкий способ построения произвольного симметричного слова длины $2k + 1$. Возьмём

произвольное слово q из k букв. Припишем к нему справа (к концу слова) произвольную букву W (не совпадающую, разумеется, с последней буквой слова q), затем перевернём слово q и припишем полученный «перевёртыш» справа. Мы получим симметричное слово.

Какое преобразование выражает слово такого вида? Как рассказывается в курсе эстетической геометрии (или любом учебнике по группам преобразований), это слово задаёт симметрию относительно $q(W)$, где q — преобразование, определённое словом q (будем пользоваться одним обозначением для слова и для определённого им пребразования), а W — окружность, обозначенная буквой W. Итак, каждое симметричное слово определяет симметрию относительно окружности. Ранее отмечалось, что окружность и симметрию относительно окружности мы обозначаем одинаково. Теперь же появляется и *слово*, которое выражает то симметрию относительно окружности, а то саму окружность, в зависимости от контекста. В дальнейшем изложении все строчные латинские буквы означают симметричные слова. А исходные окружности, названные буквами алфавита, по-прежнему будут обозначаться прописными латинскими буквами.

На рис. 9 мы видим часть фрактального калейдоскопа, созданного тремя окружностями A, B, C (сплошной, штриховой, пунктирной). Исходные окружности представляют слова длины 1: A, B, C. Следующие по уменьшению окружности — слова длины 3: ABA, ACA, BAB и т. п. — всего шесть таких слов. Ещё меньшие окружности — слова длины 5, например: $ABCBA$ или $CBCBC$. Самые маленькие окружности — слова длины 7, как $ABACABA$ или $CBACABC$.

Как, глядя на окружность, прочитать слово, которое она означает? Посмотрим, скажем, на маленькую сплошную окружность внутри штриховой окружности B. Всякая окружность, лежащая внутри B, получена симметрией относительно B, потому первая (и последняя) буква в этом слове — B. Находим образ этой маленькой окружности после симметрии относительно B. Это сплошная окружность внутри окружности A. Всякая окружность, лежащая в A, получена симметрией относительно A. Следовательно, вторая (и предпоследняя) буква интересующего нас слова — это A. После симметрии относительно A мы получим сплошную окружность уже побольше, лежащую в B, так что третья с начала (и с конца) буква B. После симметрии полученной окружности относительно B у нас будет окружность A. Поэтому в центре слова написана буква A. Само же искомое слово имеет вид $BABABAB$. Аналогично можно прочитать любую изображённую окружность. Каждый раз процесс заканчивается, когда после симметрии мы попадаем в одну из исходных окружностей алфавита.

Слова длины 1 — это исходные окружности A, B, C.

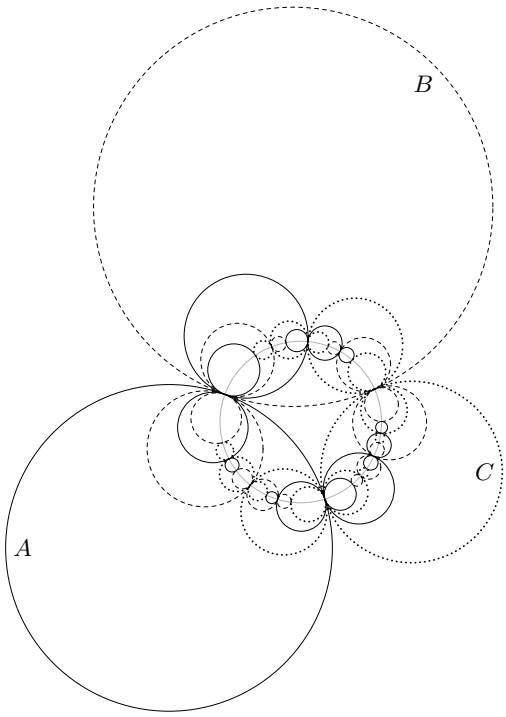


Рис. 9

Чем длиннее слово, тем меньше представляющая его окружность, и тем ближе она подходит к окружности I, перпендикулярной исходным окружностям A, B, C. Все окружности, представленные симметричными словами, никогда не пересекаются. Поэтому слово f (точнее, «окружность, представленная словом f», но для краткости мы будем это опускать) обязательно разделяет слово h и слово fhf (так как эти слова представляют окружности h и $f(h)$, т. е. пару окружностей, симметричную относительно f). Мы выбрали окружности алфавита так, что ни одна окружность не лежит внутри другой. Поэтому если X — какая-то буква алфавита, то любое симметричное слово, начинающееся с буквы X, лежит внутри X (это слово представляет результат симметрии относительно X). Мы обсуждали подобный вопрос ещё до того, как ввели понятие алфавита: надо было убедиться, что разные слова представляют разные преобразования. Пусть симметричное слово f начинается с букв XY, т. е. имеет вид $XYhYX$. Представленную им окружность можно выразить композицией симметрий: $f = X(Y(h))$. Так как YhY лежит внутри Y, окружность f лежит внутри XYX. Аналогично слово $XYZhZYX$ лежит внутри не только внутри X и XYX, но и внутри $XYZYX$. Таким образом, чтобы понять, внутри каких

окружностей лежит слово f, нам не всегда нужно дочитывать f до конца (точнее, до середины, потому что f симметрично). Как в десятичной записи числа слева стоят бóльшие разряды и мы можем приближённо оценить число, не обращая внимания на стоящие справа цифры (число тысяч нам важней числа единиц или десятков), так можно примерно оценить расположение окружности f, не дочитывая слово f до конца. В общем случае, если слово f имеет вид $hXgXh$ (где g — некое симметричное слово меньшей длины), а X — буква алфавита, то f лежит внутри слова hXh, где h — некое не обязательно симметричное слово, с которого начинается слово f.

Рассмотрим простейший случай: алфавит состоит всего из двух букв A, B. Тогда любое слово состоит из последовательности чередующихся букв A и B. Все слова можно разделить на два множества: в одном множестве слова, начинающиеся с A, во втором — начинающиеся с B. Все слова нечётной длины оказываются симметричны, поэтому каждое слово нечётной длины в алфавите из двух букв представляет собой окружность. Как устроены эти окружности? Они показаны на рис. 3, иллюстрирующем понятие *пучка*, поскольку все они лежат в одном пучке, в котором лежат A и B. Рассмотрим все слова нечётной длины, начинающиеся с A: A, ABA, $ABABA$. Эти слова представляют череду вложенных друг в друга окружностей. Существует только одна точка, которая лежит *внутри всех* этих окружностей. Эта точка называется «центром пучка окружностей» Теперь рассмотрим слова нечётной длины, начинающиеся с B: B, BAB, $BABABAB$, … Этот ряд слов-окружностей совершенно аналогичен уже рассмотренному. Существует только одна точка, лежащая внутри всех рассматриваемых окружностей. Эта точка будет вторым центром пучка окружностей. Заметим одно важное для нас исключение. Если окружности A и B *касаются* друг друга, а именно с такого случая мы и начали наше рассмотрение, то центры пучка совпадают друг с другом и с точкой касания этих окружностей. Эта точка лежит не внутри рассматриваемой последовательности окружностей, а на всех этих окружностях. Введённое раньше понятие *изгиба* окружностей даёт численную меру для всех симметричных слов в этом случае. Каждому слову, симметричному слову-окружности, приписывается (A–B) координата; слова нечётной длины, начинающиеся с буквы A, имеют координаты $0, -1, -2, -3, \ldots$, а слова нечётной длины, начинающиеся с буквы B, имеют (A–B) координаты $1, 2, 3, \ldots$ Этот случай можно увидеть на рис. 9, если рассматривать только сплошные и штриховые окружности.

Между двумя рассмотренными последовательностями слов-окружностей есть простое соответствие. Заменим в каждом слове единовременно букву A на букву B, а букву B на A. Тогда A соответствует B, ABA

соответствует BAB, и так далее. Какой *геометрический смысл* этого соответствия? Произведённая замена букв наглядно изображается симметрией относительно биссектрисы (серединной окружности) между A и B. При этой симметрии меняются местами окружности A и B и все окружности двух разобранных последовательностей. Также меняются местами центры пучка. А что происходит с (A–B) координатами (для случая, когда A и B касаются друг друга)?

Можно сказать, хотя это и не очень строго: чем длиннее симметричное слово, тем меньше представленная им окружность. Более того, если мы рассмотрим симметричные слова длины n, то с увеличением n совокупный объём этих слов-окружностей становится всё меньше, стремясь к нулю, хотя самих окружностей становится всё больше и больше. Если окружностей алфавита всего три: A, B, C, то обязательно существует окружность I, перпендикулярная им всем. Окружности, представленные симметричными словами из этого алфавита, также обязательно будут перпендикулярны I. Это уже помогает нам представить расположение окружностей.

Что происходит, если A, B, C касаются друг друга? Тогда I проходит через точки их касания, а окружности A, B, C накрывают I полностью, всякая точка I лежит внутри какой-то из этих трёх окружностей. Ко всякой точке x, лежащей на I, можно приблизиться охватывающими её симметричными словами-окружностями. Это делается так: x лежит либо внутри A, либо внутри B, либо внутри C. Пусть это окружность A. Слова-окружности ABA и ACA покрывают целиком ту часть I, которая лежит в A, следовательно, x лежит внутри одной из этих окружностей. И так далее. Мы получим последовательность слов-окружностей, стягивающихся к произвольной точке x. Это напоминает запись точки в десятичной (или троичной) системе счисления. В десятичной системе счисления разные последовательности знаков выражают разные числа, за некоторыми исключениями. Десятичная дробь 0,99999999 . . . , где девяток бесконечно много, означает то же, что и число 1. Подобное случается и когда мы рассматриваем бесконечное симметричное окружность-слово. Это связано с точками касания исходных A, B, C между собой. Будем рассматривать слова из двух букв A и B. Слово AB означает преобразование, стягивающее все точки плоскости к s — точке касания A и B. Преобразование $ABAB$ — стягивает к s в два раза быстрее, $ABABAB$ стягивает в три раза быстрее, и так далее. Преобразования, начинающиеся с буквы B: BA, $BABA$, $BABABA$ также стягивают всё к точке s (но с другой стороны). Поэтому преобразование-слово чётной бесконечной длины стягивает всю плоскость в точку s.

Разумеется, эту мысль можно и нужно уточнять. Мы рассмотрим симметричные слова-окружности из двух букв. Любые две такие окружности

касаются друг друга в точке s, причём окружности-слова, начинающиеся с A, касаются с одной стороны, а начинающиеся с B — с другой. С увеличением длины такого слова оно означает окружность всё меньшего размера, которая становится малоотличима от точки s. Таким образом, к точке s можно приблизиться *двумя* последовательностями слов-окружностей. Аналогичными свойствами обладают и две другие точки касания p и q. И если мы будем отражать любую из точек касания относительно трёх окружностей алфавита, то к получаемым точкам также можно приближаться с разных сторон. Семейство точек, полученных из исходных точек касания симметриями относительно окружностей алфавита, постепенно распространяется по всей окружности I. Они не заполняют целиком никакой дуги окружности I, но на произвольной, сколь угодно малой дуге окружности I обязательно найдутся точки из этого семейства, подобно тому как рациональные числа есть внутри сколь угодно малого отрезка числовой прямой.

Теперь рассмотрим случай, когда три исходные окружности алфавита A, B, C не имеют общих точек. Тогда перпендикулярная им окружность I *не полностью* накрывается окружностями $A, B, C, \ldots$ Что происходит с симметриями трёх исходных окружностей друг относительно друга? Они выражаются симметричными словами-окружностями. Все симметричные слова длины 3 (их шесть, внутри каждой исходной окружности лежат 2 другие) накрывают симметричные слова длины 5 (снова внутри каждой окружности длины 3 лежит 2 окружности длины 5, всего таких окружностей 12) и так далее: с увеличением длины слова на 2 число симметричных слов данной длины удваивается. При этом каждая новая окружность перпендикулярна I. Мы можем отмечать на I «дыры». Окружности-слова длины 1 не накрывают все I. Окружности-слова длины 3 накрывают ещё меньшую часть I: внутри дуги I, накрытой окружностью A, появляются дыры. Окружности-слова длины 5 накрывают ещё меньшую часть, образуя новые дыры. И так далее. Постепенно «дыр» становится всё больше и их совокупная длина оказывается всё ближе к длине всей окружности I. Соответственно, длина дуг окружности I, накрытых симметричными словами-окружностями длины k, становится всё меньше и меньше, стремясь к нулю. В отличие от случая, когда A, B, C касались друг друга, вовсе не к каждой точке, лежащей на I, удастся сколь угодно близко подобраться с помощью симметричных слов-окружностей длины k. Наоборот, вероятность того, что к точке x можно так подобраться, равна нулю.

Что получается в пределе, если отражать окружности бесконечно долго? Они стянутся в некоторые точки, лежащие на I. Легко понять, что этих точек бесконечно много, но они лежат «порознь» друг от друга. Похожими

свойствами обладает, например, множество чисел на отрезке от 0 до 1, в десятичной записи которых нет цифры 9. А множество «дыр» напоминает, наоборот, множество таких чисел, в десятичной записи которых есть цифра 9. Хотя на первый взгляд чисел «без девятки» куда больше, чем чисел «с девяткой», на самом деле вероятность того, что среди десятичной записи произвольной точки нет девятки (ни в одном разряде!) — равна нулю.

Совокупность точек, к которым стягиваются все окружности после бесконечных отражений друг относительно друга, обозначим W. Все эти точки, напомним, лежат на I.

Теперь посмотрим на все слова в алфавите A, B, C. Эти слова означают некоторые преобразования плоскости. Каждое из этих преобразований имеет предельную точку (куда стягивается вся плоскость при многократном применении этого преобразования). Эта предельная точка обязательно лежит в W. Чем длиннее слово, тем быстрее означаемое им преобразование стягивает всё к предельной точке, если же слово бесконечной длины, то оно отображает все точки плоскости сразу в одну, лежащую в W. Мы можем так и мыслить W — как совокупность точек плоскости, куда обязательно попадёт любая точка плоскости под действием какого-нибудь слова бесконечной длины (отражаясь достаточно долго от трёх окружностей алфавита). Это множество W принято называть «фрактальной пылью». Во фрактальную пыль стирает всю плоскость действие бесконечных слов-преобразований.

Пока мы анализировали свойства окружностей, возникающих из симметрий относительно нескольких данных окружностей алфавита, *качественно*. Хочется ввести какую-то количественную меру, позволяющую записывать точные равенства и уравнения. Этим, используя определённый ранее *изгиб* окружностей, мы и займёмся в следующем параграфе.

§ 6. Симметрии относительно трёх окружностей в (A–B) координатах. Дробно-линейные преобразования и цепные дроби. Уравнение золотого сечения

Как мы видели, понятие изгиба или (A–B) координат даёт возможность ввести координаты для любой точки плоскости. Если у нас есть три касающиеся окружности A, B, C, то любая точка окружности I, перпендикулярной им всем, *однозначно* задаётся своей (A–B) координатой, а также своей (B–C) или (C–A) координатой. Правила перехода от одной системы координат к другой мы свели в таблицы. Причём были указаны два варианта введения координат: с помощью окружностей A, B, C и с помощью точек касания этих окружностей: p, q, s. Теперь мы изучим, как меняются координаты точек на I под воздействием симметрий относительно A,

B и C. Это позволит нам описать, как действуют слова-преобразования, составленные из алфавита A, B, C, на точки окружности I.

Начнём с простейшего. Пусть x — и точка, и её (A–B) координата на окружности I. Чему равна (A–B) координата точки $A(x)$? Легко видеть, она равна $-x$. Чему равна (A–B) координата точки $B(x)$? Также нетрудно видеть, она равна $2 - x$. Теперь мы легко увидим, как меняются координаты точки x при композиции симметрий относительно A и B: $A(B(x)) = -(2 - x) = x - 2$; $B(A(x)) = x + 2$. Таким образом, точка с бесконечной координатой остаётся на месте, а остальные точки сдвигаются на 2 в ту или другую сторону. Что означают слова, составленные из двух букв A и B? Слово чётной длины $2k$, начинающееся с A, означает функцию $x - 2k$, начинающееся с B — функцию $x + 2k$. Слово нечётной длины при двухбуквенном алфавите всегда означает симметрию относительно некоторой окружности и потому должно выражаться инволютивной функцией. Установив, как работают симметрии относительно A и B и какие функции задают слова, составленные из этих двух букв, перейдём к симметрии относительно окружности C. Чему равно $C(x)$? Ответ не очевиден. Воспользуемся следующим приёмом: мы легко можем выразить эту зависимость в (A–C) координатах. Это точно такая же зависимость, как для $B(x)$ в (A–B) координатах. Поэтому найдём (A–C) координаты точки x, это нами обозначено было как $f(x)$, затем найдём $B(f(x))$, а полученное выразим в (A–B) координатах, результат будет $C(x) = f^{-1}(B(f(x)))$. Мы знаем, что $f(x) = 1/x$, $B(f(x)) = 2 - f(x) = 2 - 1/x$, подставим: $f^{-1}(B(f(x))) = 1/(2 - 1/x) = x/(2x - 1)$. Мы сталкиваемся с подобным приёмом, не только работая с координатами или системами отсчёта, но и в более разнообразных ситуациях, например при переводе. Общаясь с англичанином, мы, чтобы узнать ответ на свой вопрос, переводим вопрос с русского на английский, а ответ переводим с английского на русский. Функция f в этом случае означает «перевод с русского на английский» а обратная ей f^{-1} — «перевод с английского на русский». При этом методе у нас неизбежно появляется понятие сопряжённых элементов или сопряжённых функций: если f — правило перехода из одной системы координат в другую, а Q — некое построение или вычисление, какое нам удобней делать в этой новой системе координат, то $H(x) = f^{-1}(Q(f(x))$ называется сопряжённым с $Q(x)$.

Мы выразили, как действуют симметрии относительно трёх данных касающихся окружностей A, B, C на точки, лежащие на I: $A(x) = -x$, $B(x) = 2 - x$, $C(x) = x/(2x - 1)$. Составляя композиции, мы будем использовать дробно-линейные функции. Правда, не все дробно-линейные функции выражаются такими словами: их коэффициенты не произвольные вещественные числа, а целые числа особого вида. Сейчас мы не будем

углубляться в изучение того, какими должны быть эти коэффициенты, а заметим связь между трёхбуквенными словами и цепными дробями.

Мы знаем, что $C(x)=x/(2x-1)$. Поделим числитель и знаменатель на x, получим $C(x) = 1/(2 - 1/x)$. Применяя к этому выражению композиции, составленные из симметрий относительно A и B, получим выражение вида $2k \pm \frac{1}{2-1/x}$. Применив к нему симметрию относительно C, получим

$$\cfrac{1}{2 - \cfrac{1}{2k \pm \cfrac{1}{2 - \cfrac{1}{x}}}}.$$

Мы видим, что получается цепная дробь. Таким образом, каждому слову из трёх букв A, B, C можно сопоставить цепную дробь, коэффициенты которой однозначно определены. Количество знаков деления в этой дроби в два раза больше, чем вхождений буквы C в слово.

После аналитического, численного выражения симметрии относительно окружности с помощью (A–B) координат появляется возможность составлять уравнения, решениями которых будут числа. До этого мы могли составлять уравнения, решениями которых были бы точка или окружность. Самое простое уравнение: $A(x) = x$. Этому условию удовлетворяют две точки пересечения A с I: q и s. Координаты этих точек — 0 и ∞. Аналогично разбираются два других уравнения. В случае $B(x) = x$ или $2 - x = x$ решения этого уравнения 1 и ∞, им соответствуют точки p и s (точки пересечения B с I). Уравнение $C(x) = x$ означает $x/(2x - 1) = x$, корни этого уравнения 0 и 1, точки q и p, по которым C пересекается с I. Мы видим, что решения уравнений согласуются с геометрией, но, разумеется, не узнаем ничего нового. Так же обстоит дело при рассмотрении уравнения $A(B(x)) = x$. У этого уравнения единственное решение: ∞ или точка s, где касаются A и B. Аналогично обстоит дело и с уравнениями $B(C(x)) = x$ или $A(C(x)) = x$. Аналитическое, численное решение лишь подтвердит очевидный геометрический факт: неподвижной остаётся только точка касания окружностей.

Первое содержательное уравнение, решение которого не очевидно геометрически, выглядит так: $C(B(A(x))) = x$. В (A–B) координатах оно записывается как $(2-(-x))/(2(2-(-x))-1) = x$. Раскрывая скобки и приводя подобные члены, получим: $(2 + x)/(2x + 3) = x$. Это уравнение приводится к квадратному $x^2 + x - 1 = 0$. Это то же самое уравнение, которое возникает при определении *золотого сечения*. Неожиданно, что первое интересное уравнение, которое можно написать, используя понятие изгиба,

имеет своим решением золотое сечение. Ведь золотое сечение возникает в геометрии прямых и казалось не связанным с геометрией окружности.

Отметим простой способ наглядно находить приближённое решение уравнения $C(B(A(x))) = x$. Возьмём произвольную точку плоскости. Отразим её последовательно относительно окружностей A, B, C в указанном порядке. С полученной точкой сделаем то же самое, и т. д. В итоге мы построим последовательность точек, сходящуюся к неподвижной точке преобразования $C * B * A$, — одному из корней изучаемого уравнения. Взяв член этой последовательности с большим номером, мы получим неплохое приближённое решение уравнения. На рис. 10 это и показано.

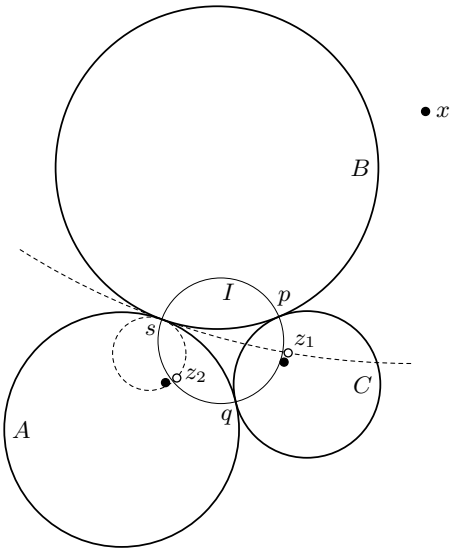


Рис. 10

Выбрана произвольная точка x, с неё начинается последовательность чёрных точек $C(B(A(x)))$, $C(B(A(C(B(A(x))))))$, ... Эта последовательность быстро прижимается к окружности I, и уже после второго шага её члены мало отличаются друг от друга. Член этой последовательности с большим номером изображён белым кружком z_1, с хорошей точностью z_1 является искомым решением уравнения $C(B(A(x))) = x$, а изгиб точки z_1 в (A–B) координатах равен золотому сечению. Штрихами указана окружность из пучка (A, B), проходящая через z_1. Уравнение $C(B(A(x))) = x$ имеет и второе решение; чтобы приблизиться к нему, нужно двигаться от x в другую сторону, то есть рассмотреть последовательность точек: x, $A(B(C(x)))$, $A(B(C(A(B(C(x))))))$, ... Её предел есть также решение уравнения $C(B(A(x))) = x$, белый кружок z_2 — член этой последовательно-

сти с большим номером — является приближённым решением уравнения. Штрихами показана окружность из пучка (A, B), проходящая через z_2. Её изгиб отрицателен, он приближённо равен второму корню полученного ранее квадратного уравнения (к первому корню близка $(A–B)$ координата точки z_1, приближённо равная золотому сечению).

Вспомним, что мы начали с трёх касающихся окружностей, и сформулируем результат так: если достаточно долго отражать произвольную точку относительно трёх касающихся окружностей (в указанном порядке), то $(A–B)$ координата точки будет мало отличаться от золотого сечения. Это — предельное свойство золотого сечения в геометрии окружности. А указанный способ приближённого решения уравнения подходит для многих уравнений геометрии окружности. И часто его графическая реализация приводит к эстетически значимым результатам. Таким образом строятся интереснейшие барочные спирали, о которых рассказывается на сайте [7]. В анализе подобный метод решения уравнений вида $f(x) = x$ называется итерационным методом Ньютона, обычно он применяется к решению уравнений в вещественных числах.

§ 7. Пары точек и гармоническое отношение

Ранее мы уже отметили, что окружность, перпендикулярная I, однозначно задаётся своими точками пересечения с I и что каждые две точки на I определяют перпендикулярную I окружность. Поэтому, как было показано в § 3, можно говорить не об окружностях, перпендикулярных I, а о парах точек, лежащих на I. Пара точек (w, v), лежащих на I, задаёт такое отображение t точек на I, что $t(x)$ есть точка, симметричная x относительно окружности Q, проходящей через w и v (см. рис. 11). Мы можем назвать это отображение сужением симметрии относительно Q на окружность I. Хотя всё действие происходит на I, геометрически (как сделать это аналитически — мы покажем позднее) построить по произвольной точке x точку $t(x)$ проще, выйдя за пределы I. Есть два способа. Первый: провести через (w, v) окружность, перпендикулярную I, а затем провести через x окружность, перпендикулярную I и новопостроенной. Вторая точка пересечения этой окружности с I (первая — исходная точка x) и будет искомой $t(x)$. Второй способ возвращает нас к миру

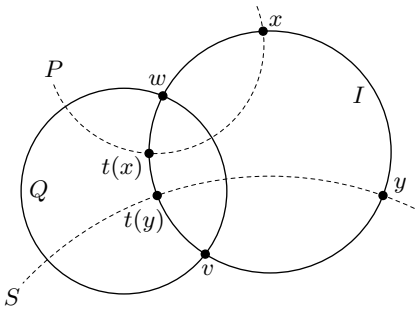


Рис. 11

прямых. Проведём через w и v касательные (см. рис. 12). Точка их пересечения A будет центром окружности, перпендикулярной I и проходящей через w и v. Проведём через A и x прямую. Вторая точка пересечения этой прямой и I (первая — исходная точка x) и будет искомой $t(x)$. Хорошее упражнение — доказать, что оба способа дают один и тот же результат.

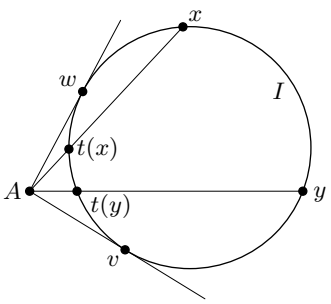


Рис. 12

Окружность Q перпендикулярна I, вспомогательные окружности P и S перпендикулярны им обеим. Разумеется, $t(t(x)) = x$, $t(t(y)) = y$. Поскольку оба способа дают один и тот же результат, можно сформулировать теорему планиметрии: выбрав произвольные точки x и y, мы получим точки $t(x)$ и $t(y)$, как показано на рис. 11, найдём A — точку пересечения соответствующих прямых, как показано на рис. 12. Теорема утверждает, что прямые, проходящие через A и w или через A и v, касаются I. Можно по-разному переформулировать эту теорему, например: прямые, проходящие через пары «образ-прообраз» отображения t, обязательно пересекаются в одной точке A.

Первый способ связан с *биплетной* симметрией плоскости: так в геометрии окружности называется симметрия плоскости относительно *пары точек*. Второй способ основан на A-отображениях. О них подробно говорится в [8]. Среди прочего там рассказывается о моделировании A-отображениями проективной геометрии. Описанное только что понятие симметрии одной пары точек относительно другой пары точек в проективной геометрии называется гармоническим расположением (или отношением) пар точек. Обратим внимание, что симметрию относительно пары точек можно мыслить как выворачивание наизнанку дуги, ограниченной этими двумя точками. Точке, лежащей по одну сторону от пары точек, симметрична точка, лежащая по другую сторону от этой пары.

Вернёмся на окружность I, где пара точек x и $t(x)$ симметрична относительно пары w, v. Свяжем эту симметрию с $(A–B)$ координатами. Как было показано в § 3, можно говорить не про $(A–B)$ координаты, а про (q, s, p) координаты, где q, s, p — точки касания исходных окружностей между собой. При этом точка q соответствует нулю, точка s соответствует бесконечности, точка p соответствует единице. Какова координата точки, симметричной s относительно p и q? Ответ прост: через 0 и 1 проходит ортогональная I окружность C, поэтому мы ищем такую точку s, что $C(s) = C(\infty) = 1/2$. Это можно получить, используя формулу $C(x) = x/(2x - 1)$. В данном случае $x = \infty$, поэтому нужно вычислить предел. Можно этот же результат

получить геометрически, используя тот факт, что окружность, проходящая через s и $C(s)$ (перпендикулярная I, как и все рассматриваемые нами окружности), будет биссектрисой между A и B, а $(A–B)$ координата биссектрисы равна $-1/2$. А как найти точку (её координату), симметричную p относительно q и s? Можно воспользоваться заменой координат (перейти к $(A–C)$ координатам), но проще заметить, что эта точка есть $A(1) = -1$. Аналогично $C(0) = 2 - 0 = 2$.

Полученные три точки также можно отражать относительно исходных окружностей. Полученные точки — снова отразить. И так далее. Ранее мы изучали этот процесс качественно в § 5, а в § 6 связали композиции относительно A, B, C с цепными дробями. Заметим, что на каждом шаге (кроме первого) точку можно отразить относительно *двух* окружностей, хотя всего окружностей *три*. Ведь мы попали в данную точку после симметрии относительно какой-то окружности, и симметрия относительно неё же вернёт нас обратно. Полученное множество точек всюду плотно распределено на окружности I.

Термин «гармоническое отношение» возник в проективной геометрии для описания двух пар точек, лежащих на одной прямой. Проективная геометрия возникла в связи с изучением перспективы в живописи: точку «схода» всех «параллельных» на картине изучают методами проективной геометрии. Точнее — изучение этой точки, понимание законов перспективы художниками подталкивало математиков к развитию проективной геометрии, расцветшей в XIX веке. Многое для неё сделал ещё Паскаль. Хотя проективная геометрия изучает прямые и гармоническое отношение определяется для точек, лежащих на прямой, а мы сейчас изучаем геометрию окружности, перейти от одного к другому несложно. От геометрии окружности перейти к геометрии прямых в данном случае просто, так как все симметрии происходят на окружности I. Достаточно поэтому рассмотреть инверсию с центром в точке, лежащей на I. Тогда I перейдёт в прямую, а точки, симметричные относительно пары точек, — снова в симметричные точки. Напоминаю подробно разобранный в учебнике тезис: симметричное симметрично симметричному. В проективной геометрии можно довольно неожиданным образом определить если не окружность, то её обобщения — кривые второго порядка. Но куда удобней глядеть из геометрии окружности на проективную геометрию, чем наоборот.

Теперь введём определение *гармонического отношения* двух пар точек на окружности: если пара точек (x, y) симметрична относительно пары точек (w, v), то мы называем пары точек (x, y) и (w, v) гармоническими парами или «гармонически разделяющими друг друга». В этом случае точки x и y обязательно лежат по разные стороны от каждой из точек w

и v, а точки w и v — по разные стороны от каждой из точек x и y. Укажем на важнейшее свойство гармонических пар: свойство *взаимности*. Если x и y симметричны (гармонически разделены) относительно w и v, то обязательно точки w и v симметричны (гармонически разделены) относительно x и y. Лучше всего это видно из первого способа построения симметричных точек, связанного, как было сказано, с биплетным исчислением. Проведём через w и v окружность J, ортогональную I (на которой лежат все наши точки), и окружность U, проходящую через x и перпендикулярную J и I. Вторая точка пересечения U и I (обозначенная как y) будет симметрична x относительно w и v. Рассмотрим теперь симметрию относительно U (на окружности I это будет совпадать с симметрией относительно пары точек x и y). Тогда $U(w) = v$, так как $U(I) = I$ и $U(J) = J$, поэтому точки пересечения I с J (w и v) меняются местами. Что и требовалось. Можно понять свойство взаимности немного иначе. Заметим, что окружности J и U абсолютно равноправны: они обе перпендикулярны друг другу и I, и свойство взаимности лишь переформулирует это геометрическое свойство.

Теперь посмотрим на гармоническое отношение с помощью (q, s, p) координат. Как мы видели, если точка x симметрична точке s относительно q и p, то её (q, s, p) координата равна $1/2$. Можно обозначить это так: $g(q, s, p, x) = 1/2$. После перестановок получим: $g(s, q, p, x) = 2$, $g(p, q, s, x) = -1$. Другие перестановки трёх точек новых значений не дадут. Заметьте, что два новых полученных значения -1 и 2 совпадают с полученными ранее значениями $A(1)$ и $B(0)$. В этом есть простой геометрический смысл. Обратим внимание, что в последней записи $g(p, q, s, x) = -1$ точки следуют именно в том порядке, в каком входят в пары.

Введённое обозначение $g(q, s, p, x)$ само подсказывает нам, что его можно истолковать более общим способом: речь идёт о функции, сопоставляющей любым четырём точкам на окружности вещественное число, равное координате последнего аргумента (точки x), исчисленной относительно первых трёх точек (взятых именно в указанном порядке). Это открывает нам возможность искать новые зависимости: теперь мы можем переставлять и четвёртую переменную x. Поэтому возникает вопрос: как, например, связаны $g(q, s, p, x)$ и $g(q, s, x, p)$? На этот и подобные вопросы мы ответим в следующем параграфе.

§ 8. Подробнее о калибровке окружности. Двойное отношение. Соотношения между изгибами

В § 3 уже говорилось о связи изгиба и смены координат с калибровкой линейки. Чтобы превратить прямую в линейку, достаточно выделить на ней

две точки A, B и написать у одной «0», а у другой «1». Введём функцию трёх переменных $k(A, B, C)$, означающую координату точки C, если 0 отмечен в точке A, а 1 — в точке B. Легко понять, что $1 - k(A, B, C) = k(B, A, C)$. Функция $k(A, B, C)$ выражается формулой

$$k(A, B, C) = \frac{|A, C|}{|A, B|},$$

где $|\ |$ означает расстояние между точками. Эта функция называется ещё «простым отношением» (трёх точек A, B, C). При этом расстояния надо считать со знаком, ориентированными: $|A, B|$ (мера длины) всегда положительно, а $|A, C|$ положительно, если C лежит на прямой по ту же сторону от A, что и B. Если же A разделяет B и C, то $|A, C|$ — отрицательно. Теперь поступим аналогично тому, как мы поступили, изучая изгиб. Будем считать $k(A, B, C)$ не только «координатой» точки C (при единице измерения $|A, B|$), но функцией трёх точек. Тогда естественно изучать не только связь между $k(A, B, C)$ и $k(B, A, C)$, но и связь между $k(A, B, C)$ и $k(A, C, B)$. Она тривиально получается из формул

$$k(A, B, C) = \frac{|A, C|}{|A, B|}, \quad k(A, C, B) = \frac{|A, B|}{|A, C|}.$$

Те же самые функции, которые управляли сменой координат изгиба окружностей, появляются и здесь. Это не очень удивительно, ведь в обоих случаях речь идёт о перестановках трёх объектов, каждой перестановке отвечает функция, управляющая значениями координат. Отсюда сразу следует инволютивность функций, управляющих переменой мест двух букв, и тройственная симметрия их композиций.

Вернёмся к рассмотрению окружности. В предыдущем разделе введена функция четырёх переменных: $g(q, s, p, x)$. Свяжем её с только что сделанными построениями на прямой. Это можно сделать, например, с помощью проектирования из точки s на рис. 8. В этом случае точка s переходит в бесконечно удалённую точку. Её мы трогать пока не будем. Посмотрим на g как на функцию *трёх* переменных q, p, x. На прямой проектирования она совпадает с рассмотренной только что функцией k. На прямой проектирования точка q' играет роль нуля (точки A), p' — роль единицы (точки B), а точка x' — роль точки C (см. рис. 8). Тогда

$$g(q, s, p, x) = k(q', p', x') = \frac{|q', x'|}{|q', p'|}.$$

Теперь мы возвращаемся к функции g, так как функция k выполнила своё первое предназначение. Ранее мы изучили и даже показали в таблицах,

как меняется функция g, если менять местами три первые её переменные q, s, p. Рассмотрев, как меняется функция k при перестановках переменных, мы понимаем, что происходит при перемене местами двух последних переменных: p и x. Именно,

$$g(q,s,x,p) = k(q',x',p') = \frac{|q',p'|}{|q',x'|} = \frac{1}{g(q,s,p,x)}.$$

При перестановке третьего и четвёртого аргумента g снова возникает функция $1/z$, так как $f(g(q,s,p,x)) = g(q,s,x,p) = g(s,q,p,x)$ (последнее равенство было установлено уже давно). Поэтому с помощью композиций функций f и h можно выразить изменение значений функции g. Обратим внимание на появившееся равенство $g(q,s,x,p) = g(s,q,p,x)$. Перемена мест сразу в двух парах аргументов не меняет значение функции g. Подобное происходит при любом выборе пар аргументов. Равенства можно доказать, рассматривая композиции, управляющие заменами аргументов. А можно доказать и геометрически.

Именно геометрический смысл этих тождеств мы сейчас разберём. Вернёмся к рис. 7. При исходном порядке аргументов $g(q,s,p,x)$ равно изгибу окружности K в пучке (A,B). Если мы поменяем местами первый с третьим аргументом, второй с четвёртым, то получим $g(p,x,q,s)$. Это будет изгиб окружности K в пучке (L,M). Формула утверждает, что эти изгибы совпадают. Более наглядно сравнение изгибов окружностей, одни из которых касаются друг друга в точке s, а другие в точке p. Изгиб окружности B в пучке (K,A) равен $g(x,s,q,p)$, снова переставим местами аргументы: первый с третьим, второй с четвёртым — и получим равенство $g(x,s,q,p) = g(q,p,x,s)$. Правая часть выражает изгиб той же самой окружности B в пучке (C,L). Аналогично изгиб окружности M в пучке (A,C) равен изгибу окружности M в пучке (L,K). Так же можно проследить связи между другими изгибами, во всех случаях взаимосвязь определяется через перестановки аргументов функции g и легко считается с помощью композиции функций f и h (т. е. $1/z$ и $1-z$).

Детальный анализ изменения значений функции g после перестановок аргументов проводится методами теории групп. Среди всех перестановок аргументов выделяются такие перестановки, которые не изменяют значения функции g. Как было сказано, это такие перестановки, которые меняют местами две пары аргументов. Такие перестановки сами образуют группу. Это нормальная подгруппа группы всех перестановок аргументов.

Вернёмся к функции $k(A,B,C)$ (по сути это просто функция g с фиксированным вторым аргументом). До сих пор мы изучали, как меняются функции k и g при перестановках аргументов. Теперь же мы рассмотрим

новые переменные. Мы уже указывали на связь функции k с переходом от одной системы мер к другой. Выполняется тождество

$$k(A, B, C) \cdot k(A, C, D) = k(A, B, D).$$

Оно равносильно тождеству

$$k(A, B, C) \cdot k(A, C, D) \cdot k(A, D, B) = 1.$$

Функция k есть «усечённая» функция g (с фиксированным вторым аргументом). Поэтому имеет место следующее тождество:

$$g(q, s, p, x) \cdot g(q, s, x, y) \cdot g(q, s, y, p) = 1.$$

Здесь первые два аргумента в каждом сомножителе неизменны, а остальные три аргумента циклически сменяют друг друга, точно так, как это было для функции k.

Геометрический смысл этого тождества иллюстрируется на рис. 13. На окружности I выделены пять точек q, s, p, y, x и проведены все возможные перпендикулярные I окружности через пары этих точек. Всего имеется десять таких окружностей. Если рассматривать внутренность I, то дуги этих окружностей образуют интересную фигуру. Окружности, имеющие общую точку, обязательно в ней касаются друг с другом. В каждой из пяти точек появляются пучки окружностей, в которых можно измерять изгибы.

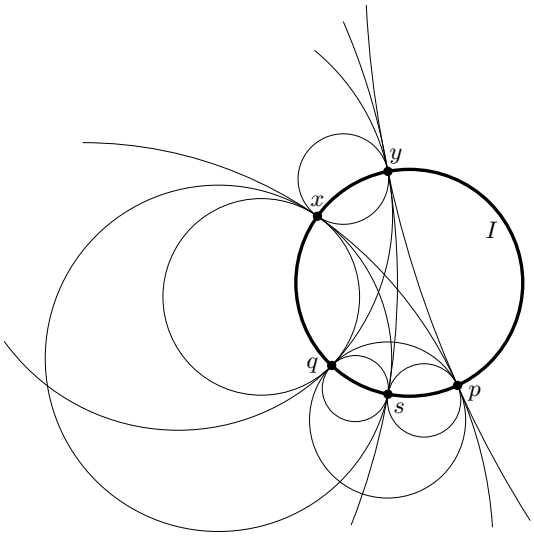


Рис. 13

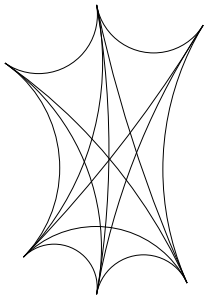


Рис. 14

Рис. 15

Интересующее нас тождество связывает изгибы окружностей, проходящих через s. Пользуясь им и правилами замены аргументов функции g, можно установить много других тождеств, связывающих изгибы построенных окружностей.

Приведём аналогичное построение, сделанное на шести точках (см. рис. 14). От всего рисунка оставлены только дуги касающихся окружностей, лежащие внутри перпендикулярной им всем I. Тут есть много интересных соотношений, но я сейчас предлагаю просто порадоваться красивой фигуре. А на рис. 15 изображена эта же фигура, но немного обработанная и повёрнутая.

Вернёмся к функции g. Мы нашли много её свойств, хочется выразить её формулой. Численное выражение функции $g(q, s, p, x)$ естественно искать в координатном виде. Пусть теперь буквы q, s, p, x означают не только точки, но и их координаты. Скажем, если $q = 0$, $s = \infty$, $p = 1$, то мы можем записать $g(0, \infty, 1, x) = x$. В общем случае ответ оказывается удивительно прост:

$$g(q, s, p, x) = \frac{(x - q)/(x - s)}{(p - q)/(p - s)}.$$

Или, используя обозначения, употреблённые в выражении для k:

$$g(a, b, c, d) = \frac{(d - a)/(d - b)}{(c - a)/(c - b)}.$$

С помощью приведённой формулы удобно проследить, как меняются значения g при перестановке переменных. Обратим внимание, что функция g является *инвариантом*: g выражает свойство четырёх точек, не зависящее от системы координат. *Инвариантность* функции g проявляется и в том, что если мы отразим четыре входящие в неё точки относительно какой-нибудь окружности — значение g на получившихся точках будет тем же самым, что и на их прообразах.

Как доказать, что функция g имеет приведённый вид? Есть много вычислительных способов, связанных с переходом от окружности I к прямой. На самом деле функция g, полагаю, первична. Выше мы указали тождество для пяти точек и функции g. Это тождество можно положить в основу определения умножения. Также, опираясь на g, можно определить и сложение, и вычитание. Здесь мы не будем этим заниматься, а в последнем параграфе укажем, как из геометрии окружности построить комплексные числа и более общие структуры. В завершение параграфа заметим: функцию g называют также *двойным отношением* (определённым на точках a, b, c, d), и эта функция широко используется в проективной геометрии. Геометрия окружности и понятие изгиба позволяют ввести его очень естественно.

§ 9. Уравнения и предельные точки. Геометрический смысл замены букв в слове

А какие вообще уравнения можно составить, оперируя симметриями относительно нескольких окружностей? Не так-то много: если Q и V — слова из нескольких букв-окружностей алфавита, то все уравнения имеют вид $Q(x) = V(x)$. Это уравнение равносильно уравнению $F(x) = x$, где $F = V^{-1}Q$. Как найти решение такого уравнения?

Прежде чем искать решение, надо уяснить, *что* мы считаем *решением* такого уравнения. Разумно считать x не произвольным объектом, а точкой или окружностью. Есть и третий вариант, качественно отличный от предыдущих: F — некое преобразование, и мы можем считать x тоже неким преобразованием. Тогда запись $F(x)$ приобретает новый смысл, её нужно прочитать как $F * x * F^{-1} = x$, что означает коммутативность преобразований F и x. Тем самым решение уравнения в этом случае сводится к поиску коммутирующих преобразований, что мы не будем разбирать, так как это уведёт от геометрии к чистой теории групп. Вернёмся к двум геометрическим случаям: x — точка или x — окружность. Прежде всего заметим, что если в алфавите всего три буквы-окружности, то окружность $x = I$, перпендикулярная им всем, является решением *любого* уравнения $F(x) = x$, так как I остаётся неподвижной при симметриях относительно всех трёх букв алфавита.

Рассмотрим случай, когда x — точка, т. е. мы ищем неподвижные точки преобразования F. Как показано в § 5, буквы-окружности алфавита задают некое множество предельных точек W, куда стягиваются все симметричные слова достаточно большой длины. Туда стягиваются и все точки плоскости. Решения любого уравнения $F(x) = x$ лежат в этом множестве. Нетрудно видеть, что к x будут стягиваться близкие (на самом деле не только близкие) точки.

Теперь нетрудно найти приближённый метод решения уравнения $F(x)=x$. Мы уже использовали его для частного случая $C(B(A(x))) = x$. Возьмём произвольную точку z и начнём на неё действовать словом F, получится последовательность точек z, $F(z)$, $F(F(z))$, $F(F(F(z)))$; эта последовательность имеет предел, и этот предел обязательно является решением уравнения $F(x) = x$. Рассмотрим уравнение $F^{-1}(x) = x$. Ясно, что решение нового уравнения будет решением старого, и наоборот. Обозначим для удобства $G = F^{-1}$ и рассмотрим последовательность z, $G(z)$, $G(G(z))$, ... Предел этой последовательности даёт ещё одно решение рассматриваемого уравнения. Итак, мы видим, что уравнение $F(x) = x$ всегда имеет решение и может иметь два решения. Возможен случай, когда оба эти решения совпадают: например, в рассмотренном ранее случае $F = AB$, $F^{-1} = BA$, где A и B — касающиеся друг друга окружности.

Мы видим некоторое сходство уравнений типа $F(x) = x$ с квадратными уравнениями: они тоже могут иметь два корня или один корень (когда оба корня совпадают). Правда, в отличие от квадратных, наши уравнения обязательно имеют корень. И мы не выяснили — могут ли они иметь *больше* двух корней? Нам станет ясно, как обстоит дело, после рассмотрения основного случая: когда алфавит состоит из трёх взаимокасающихся окружностей A, B, C. В этом случае легко показать, какую функцию выражает то или иное слово трёхбуквенного алфавита. Эти функции суть композиции симметрий относительно A, B, C, а эти симметрии выражены формулами в § 6. Их композиции в (A–B) координатах, как показано там же, обязательно имеют вид дробно-линейного преобразования:

$$F(x) = \frac{Mx + K}{Nx + T},$$

где M, K, N, T — какие-то вещественные (в нашем случае — целые, но для дальнейшего это неважно) коэффициенты. Наше уравнение сводится к виду $x = (Mx + K)/(Nx + T)$; домножив обе его части на знаменатель дроби, получим квадратное уравнение. Итак, в данном случае всякое уравнение сводится к квадратному, которое мы легко решаем.

В этом рассуждении есть неточность. Ведь мы свели слово-преобразование F к дробно-линейной функции только для того случая, когда точка x лежит на I — окружности, перпендикулярной A, B, C. Но *все* решения этого уравнения обязательно лежат на I, ведь, как было показано ранее, решения такого уравнения обязательно лежат в предельном множестве W, куда стягиваются при многократном действии симметрий относительно окружностей — букв алфавита все точки плоскости. А в нашем случае W и есть окружность I.

Мы показали в § 6, что действие всякого слова в данном трёхбуквенном алфавите на точки окружности I выражается численно с помощью дробно-линейной функции. Оказывается, почти то же самое верно в гораздо более общем случае. Если точка не лежит на окружности I, то всякое слово также будет действовать на неё как некоторое дробно-линейное преобразование. Но — с комплексными коэффициентами. Также и сама точка плоскости выражается комплексным числом. На самом деле всякое слово в *любом* алфавите, оказывается, можно представить дробно-линейным преобразованием (считая точку плоскости комплексным числом). Мы здесь не будем в это углубляться, связь комплексных чисел, вообще структуры *поля* и кватернионов с исчислением симметрий и биплетным исчислением — тема отдельного разговора.

Сведение слов-преобразований к дробно-линейным преобразованиям полезно, но при этом теряются некоторые важные идеи. Одной из них мы сейчас коснёмся. Мы рассматривали уравнение $C(B(A(x))) = x$ и нашли его решение в (A–B) координатах. С приведённым уравнением связано ещё несколько подобных ему: $A(B(C(x))) = x$ или $B(A(C(x))) = x$ — каждый раз мы как-то переставляем буквы A, B, C. Достаточно решить одно из них — то, с которого мы начали, и с помощью замены координат найти решение требуемого. Пусть теперь у нас есть произвольное уравнение $F(x) = x$, где F — некое слово трёхбуквенного алфавита. Три буквы алфавита можно менять шестью перестановками, и каждая перестановка букв определяет (индуцирует) изменение слова F. Таким образом, из F мы получим некое слово F'. Назовём полученное слово F' сопряжённым с F (с помощью данной перестановки). Тогда и F' также сопряжено с F (с помощью обратной перестановки). Ещё одно свойство сопряжения слов — транзитивность. Если первое слово сопряжено со вторым, а второе с третьим, то первое также сопряжено с третьим при помощи композиции подстановок, сопрягающих первое со вторым, а второе — с третьим. Уравнение $F'(x) = x$ будет называться сопряжённым с уравнением $F(x) = x$.

Верно и более общее утверждение. Если у нас есть произвольные слова $F, G, H, \ldots$ и мы можем про эти слова, про составленные из них уравнения что-то сказать, тогда то же самое можно сказать и про сопряжённые с ними слова F', G', H' (разумеется, сопрягать все слова должна одна и та же перестановка букв алфавита).

Теперь мы воспользуемся равноправностью исходных окружностей и геометрически проиллюстрируем сопряжение слов. Перестановка, меняющая буквы A и B, геометрически реализуется симметрией, меняющей местами окружности A и B (и оставляющей окружность C на своём месте). Это — симметрия относительно серединной окружности или биссектрисы

между A и B. Как показано в учебнике [7], окружность C обязательно перпендикулярна этой биссектрисе. Её легко описать численно с помощью (A–B) координат: она пересекает I в точке касания A и B, то есть в точке s, имеющей координату ∞, и в точке с координатой $1/2$, которая симметрична ∞ относительно пары точек 0, 1. Перестановка, меняющая местами окружности B и C, геометрически реализуется биссектрисой между B и C, при симметрии относительно этой биссектрисы окружность A остаётся на месте. Перестановка, меняющая местами A и C, реализуется биссектрисой между ними. Как и в первом случае, координаты точек пересечения этих биссектрис с I легко вычисляются. Пользуясь симметриями относительно этих трёх биссектрис, мы можем как угодно переставлять буквы-окружности A, B, C на чертеже или в слове-преобразовании.

Обозначим биссектрису между A и B буквой D: $D(A) = B$. Тогда результат перестановки букв A и B в произвольном слове w можно выразить как DwD. Аналогично можно поступить и с перестановками, меняющими местами другие буквы. Тем самым абстрактная операция замены букв в слове моделируется геометрическими преобразованиями — симметрией между окружностями алфавита. Мы могли бы просто ввести в алфавит букву-окружность D и ничего не говорить о перестановках букв A и B — это обеспечивалось действием симметрии относительно D, чисто геометрически. Правда, ранее требовалось, чтобы окружности алфавита не пересекали друг друга и были определённым образом расположены на плоскости. Это было нужно, чтобы слова, составленные из разных букв, означали разные преобразования. Включение в алфавит окружности D нарушает это требование: D пересекает C, разделяет A и B — то и другое было запрещено для окружностей алфавита. После нарушения запрета стали появляться и равенства: теперь есть разные слова, означающие одно и то же преобразование, например $DAD = B$ просто в силу определения биссектрисы. Аналогично обстоит дело с двумя другими биссектрисами-перестановками (между B и C и между A и C).

Замена букв в слове — чисто формальная операция. Возможность геометрически её истолковать и проиллюстрировать появляется потому, что любая из окружностей алфавита перпендикулярна биссектрисе между двумя другими. Если бы исходные окружности этим свойством не обладали (а, например, окружности A, $B(A)$, $B(C)$ этим свойством не обладают и ничто не мешает выбрать именно эти окружности в качестве алфавита) — то перестановки букв в словах не имели бы геометрического истолкования. Заметим, что *четыре* окружности на *плоскости* могут быть расположены таким образом, чтобы все они были «равноправны и одинаково расположены». Таковы четыре сокасающиеся окружности. А вот пять окружностей

на плоскости не могут быть «равноправны и одинаково расположены». Поэтому если мы хотим геометрически рассматривать слова пятибуквенного алфавита и перестановки букв в них — нам надо перейти в трёхмерное пространство и рассматривать пятибуквенный алфавит, составленный из сфер. Четыре сокасательные сферы также обладают нужным свойством: они «равноправны и одинаково расположены». Сфера, относительно которой симметричны две сферы-буквы такого алфавита, обязательно перпендикулярна двум оставшимся сферам-буквам алфавита. Поэтому все перестановки четырёх или пяти сокасающихся сфер-букв можно рассматривать как композиции симметрий относительно сфер-биссектрис.

§ 10. Поиск закономерностей в знаковых системах. Инверсионная диаграмма

Теперь мы применим концепцию алфавита из букв-окружностей и слов-преобразований к моделированию и поиску закономерностей. Азбука Морзе построена всего на двух сигналах: точке и тире, коротком и длинном. Но в ней, как и на письме, встречаются два одинаковых символа, стоящих подряд. Если в исходный набор (алфавит) входит три или более символа (буквы), то можно соблюсти ограничение, накладываемое геометрией окружности: два одинаковых символа не могут стоять рядом.

Пусть у нас есть какое-то устройство, испускающее в каждый момент времени один из трёх данных сигналов (позднее мы увидим, что те же методы годятся и для случая, когда исходных сигналов не три, а сколь угодно много). Под «сигналом» мы можем понимать всё что угодно: например, скрипач играет на скрипке и мы записываем нотами его музыку. Заметим, что если скрипач играет непрерывно, то два одинаковых звука (ноты) не могут стоять в записи одновременно: звук ноты «до» сольётся со звуком следующей ноты «до». Или сигналы подаёт оператор, нажимая на одну из трёх (или большее число) кнопок, а мы фиксируем, какую кнопку в данный момент он нажимает. В этом случае ещё ясней, что запрет одному и тому же сигналу стоять в записи два раза подряд имеет смысл. На клавиатуре мы, конечно, можем дважды подряд нажать на букву «А» и получить «АА». Но ведь перед тем, как нажать второй раз на букву «А», мы должны её отпустить, а это самостоятельное действие, для выражения которого надо бы создать специальный знак (сигнал).

Пусть у нас имеется три возможных сигнала, обозначим их, как обычно: A, B, C. Тогда всё происходящее можно представлять двояко. Первый, простейший вариант: в каждый момент мы регистрируем один какой-то сигнал из трёх. Тогда вся запись происходящего имеет примерно такой вид:

$A, B, C, A, C, B, \ldots$ — последовательность букв A, B, C, в которой никакая буква не записана два раза подряд. Второй вариант нам по сути не понадобится. Но метод, который будет предложен для первого варианта, годится и в этом случае (в слегка изменённом виде). Во втором варианте в каждый момент времени мы получаем не одну букву, а целое слово трёхбуквенного алфавита.

Чтобы моделировать происходящее методами эстетической геометрии, возникающие слова удобней писать «справа налево». То есть сигнал-символ, полученный первым, писать справа, слева к нему приписывать следующий символ, и так далее. В итоге сигнал, пришедший первым, читается последним, а всё слово переворачивается. В рамках нашей модели мы представим каждый сигнал-букву окружностью алфавита, а каждое слово после этого окажется зримо представленным *точкой* на плоскости. Наша цель — установить закономерности, описывающие возникающее слово.

Простейший случай закономерности — периодичность. Как из хаоса случайных слов трёхбуквенного алфавита научиться быстро выделять слова, обладающие периодичностью? Этот вопрос легко формализовать, а геометрия окружности даёт и простой ответ. Слово q называется периодическим, если оно представимо в виде $vvv\ldots = q$, где v — какое-то слово меньшей длины. Это означает, что слово q получается просто многократным повторением одного и того же слова v.

Пусть наши три буквы-окружности алфавита A, B, C — три сокасающиеся окружности. На самом деле то, что они касаются друг друга, не очень существенно, но это упрощает дело, к тому же данный случай подробно разобран. Возьмём для начала произвольную точку x, лежащую вне всех окружностей алфавита, и построим, исходя из слова, получающегося при записи поступающих сигналов, последовательность точек следующим естественным образом: «нулевая» точка последовательности — это сама точка x, первая точка — $A(x)$, вторая $B(A(x))$, третья $C(B(A(x)))$, четвёртая $A(C(B(A(x))))$, и так далее. После фиксации каждого сигнала A, B, C мы отражаем точку относительно соответствующей окружности. Таким образом, всякое слово (последовательность сигналов) означает не только композицию симметрий, но и результат действия этой композиции на точку x. В данном случае мы изобразили слово $\ldots BCACBA$, которое уже приводилось в пример. Назовём это построение *Инверсионной диаграммой* слова $\ldots BCACBA$.

На рис. 16 мы видим инверсионную диаграмму слова $CABCBA$. Видно, что уже под действием слова длины три CBA точка x очень близка к I, перпендикулярной A, B, C. Следующие точки инверсионной диаграммы будут ещё ближе к I. Внутри каждой окружности алфавита лежит по две

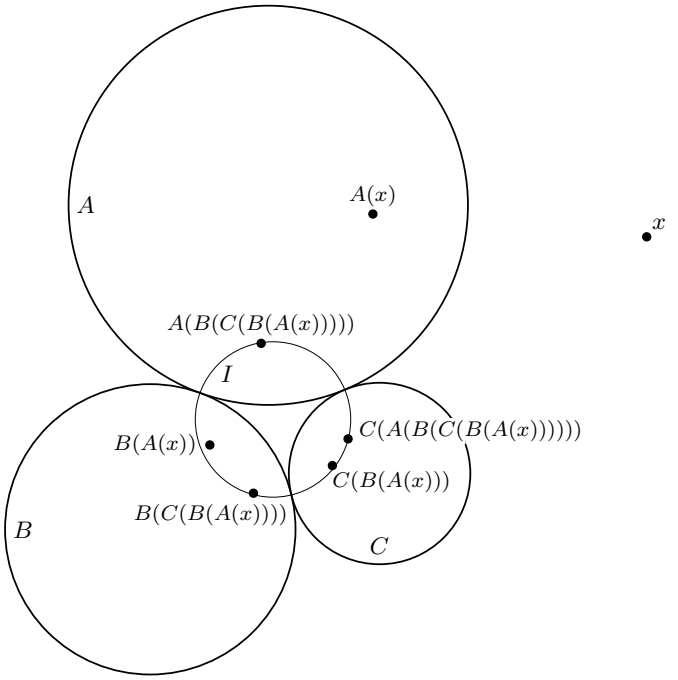


Рис. 16

точки инверсионной диаграммы, и каждая буква входит в исследуемое слово по два раза. Никаких областей сгущения на диаграмме не наблюдается (кроме неизбежного приближения точек к I).

Поясним определение *инверсионной диаграммы*. Пусть у нас есть слово q. Его инверсионной диаграммой называется последовательность точек, полученных из исходной точки x композицией симметрий относительно букв q. Композиция осуществляется в том порядке, в каком эти буквы-окружности расположены в слове q. Обратим внимание на два обстоятельства: исходная точка x выбирается почти произвольно (важно только, чтобы она была вне окружностей алфавита), следует указать направление, в котором читается слово q (справа налево или слева направо), обычно это ясно из контекста. Каждая точка диаграммы представляет «подслово» слова q, т. е. часть слова, прочитанную от его начала. Обозначим подслово длины k через q_k.

Расположение точек диаграммы слова q сообщает информацию о самом слове q и о его подсловах. Чтобы считывать эту информацию, надо воспользоваться соображениями, изложенными в § 5, где анализируются качественные характеристики слов-преобразований. Диаграмма сразу сообщает нам, сколько раз в слово q входит та или иная буква алфавита:

столько точек на диаграмме лежит внутри соответствующей окружности алфавита. В самом деле, буква входит в слово столько раз, сколько существует подслов, начинающихся с неё. А всякая точка диаграммы, представляющая подслово, начинающееся с данной буквы, — лежит внутри данной буквы-окружности. Что и требовалось.

Мы можем усложнить задачу: сколько раз в данное слово q входит какое-то слово меньшей длины q'? Начнём со случая, когда слово q' симметрично, то есть представляет собой некоторую окружность. Эта окружность легко может быть получена композицией симметрий. Количество точек диаграммы, лежащих внутри неё, и даёт ответ на вопрос. Пусть слово q' не симметрично. Внутри каких слов-окружностей обязательно лежит точка вида $q's(x)$, где s — произвольное слово? Пусть r — последняя буква слова q', $q' = q''r$ (где r, разумеется, есть A, B или C). Всякая точка y (лежащая вне r) под воздействием слова-преобразования $q''r$ перейдёт в точку, лежащую внутри $q''(r)$. Эта окружность выражается словом $q''rq''$. Поэтому и точка $s(x)$ лежит внутри этой окружности, и в ней под действием слова $q = q's$ окажется x. Итак, для того чтобы узнать, сколько раз слово q' входит в слово q, достаточно изобразить окружность $q''rq'' = q''(r)$ и подсчитать, сколько точек инверсионной диаграммы слова q лежит внутри неё. Изобразить её просто, эта окружность есть результат симметрий окружностей алфавита A, B, C друг относительно друга.

Обычно нам интересно применять инверсионную диаграмму, если слово q формируется длинной последовательностью входящих сигналов. Представим себя оператором, следящим за дисплеем, где динамически отображается инверсионная диаграмма: после приёма нового сигнала сразу появляется новая точка на диаграмме. В этом случае места, где точки расположены гуще, сразу становятся видны, и мы легко определим, внутри каких слов-окружностей эти места лежат. А это определяет, какие буквы алфавита или какие комбинации этих букв чаще входят в длинное слово q, чем другие буквы или комбинации.

Используем теперь инверсионную диаграмму для поиска периодичности в q. Для этого снова надо присмотреться к густоте точек на диаграмме. Пусть слово q периодично, т. е. имеет вид $q = www\ldots w$ (n раз). Тогда $q = w^n$ и $q(x) = w^n(x)$. В инверсионную диаграмму слова q входят точки x, $w(x)$, $ww(x)$, $www(x)$ и так далее. Эти точки сгущаются к неподвижной точке слова-преобразования w. Мы можем найти эту точку, решив уравнение $w(x) = x$. Как решать такие уравнения, разобрано в предыдущем параграфе. Фактически мы не знаем слово w, мы знаем длинное слово q и ищем w. Мы даже не знаем, существует ли w (т. е. периодично ли слово q), мы хотим иметь способ быстро узнавать, есть ли периодичность, и если есть — каков

период (т. е. найти слово w). Мы установили, что если периодичность есть, то на инверсионной диаграмме обязательно есть точка сгущения (корень уравнения $w(x)=x$). Это — не единственная точка сгущения, будет ещё $k-1$ точка сгущения, где k — длина слова. Обозначим искомое решение z, тогда точки $w_1(z)$, $w_2(z)$, $w_3(z)$, ..., $w_{k-1}(z)$, где все w_i — подслова в w, также будут точками сгущения на инверсионной диаграмме слова q, а других точек сгущения у периодического слова быть не может. Обнаружив их на инверсионной диаграмме, можно определить и само слово w. Мы можем представить себе не слово q, а последовательность сигналов, поступающих оператору. Все сигналы отображаются на экране с помощью инверсионной диаграммы. Если сигналов очень много и меняются они быстро, то видеть изображение всех сигналов сразу неудобно, мы можем увидеть лишь «грязь». Что мы увидим, если изображается только последний сигнал? Если последовательность сигналов периодическая, мы будем видеть непрерывно мигающие точки экрана в k местах — описанных ранее точках сгущения.

Мы рассматривали инверсионную диаграмму в случае, когда алфавит (базовые сигналы) состоит из трёх букв. Если букв-сигналов больше, можно поступить двумя способами. Можно промоделировать трёхбуквенным алфавитом алфавит с любым числом букв (как с помощью двух символов морзянки можно выразить любую букву любого алфавита). А можно увеличить число букв-окружностей в алфавите, чтобы каждый сигнал соответствовал какой-то букве. Если в алфавите четыре и более окружностей, то инверсионная диаграмма может выглядеть очень интересно. Ранее все точки диаграммы обязательно прижимались к окружности I (перпендикулярной всем трём окружностям алфавита), теперь уже нет никакой окружности, перпендикулярной всем окружностям алфавита, и инверсионная диаграмма может составить интересный рисунок фрактального типа. Тема эта столь широка, что мы не будем в неё входить. Заметим лишь, что важные свойства инверсионной диаграммы *помехоустойчивы*. Это означает, что если при приёме сигналов оператор напутал и внёс сигнал-букву в слово неправильно, то это локальное искажение не повлияет на общие выводы.

§ 11. Подведение итогов и пути развития эстетической геометрии

Здесь мы обсудим не только перспективы эстетической геометрии, но и её связи с другими разделами математики (напомню тезис из предисловия: «всякое математическое понятие может содержательно иллюстрироваться в эстетической геометрии»). И даже возможные связи эстетической геометрии с другими науками. Начнём с таблицы, собирающей воедино

некоторые результаты вычислений изгибов. В таблице — базовые зависимости. Окружности задаются двумя точками (точками пересечения с I). Всего есть шесть координат, но здесь мы приводим лишь три, остальные три можно получить перестановкой окружностей, с помощью функции $h(x) = 1 - x$.

№	Объект	Базис: окружности A, B, C		
		Координата A–B	Координата C–A	Координата B–C
1	Точка x на окружности I	x	$1-\frac{1}{x}$	$\frac{1}{1-x}$
2	Окружность A	∞, 0	1, ∞	0, 1
3	Окружность B	∞, 1	1, 0	0, ∞
4	Окружность C	1, 0	0, ∞	∞, 1
5	Точка касания C и A	0	∞	1
6	Точка касания B и C	1	0	∞
7	Точка касания A и B	∞	1	0
8	Произвольная окружность H, ортогональная I	k_1, k_2	$1-\frac{1}{k_1}$, $1-\frac{1}{k_2}$	$\frac{1}{1-k_1}$, $\frac{1}{1-k_2}$
9	$H(x)$	$\frac{2k_1k_2-x(k_1+k_2)}{k_1+k_2-2x}$		
10	$A(x)$	$-x$	$1+\frac{1}{x}$	$\frac{1}{x+1}$
11	$B(x)$	$2-x$	$1-\frac{1}{2-x}$	$\frac{1}{x-1}$
12	$C(x)$	$\frac{x}{2x-1}$	$1+\frac{1}{x}$	$2+\frac{1}{x-1}$
13	Окружность W пучка (A,B)	∞, k	1, $1-\frac{1}{k}$	0, $\frac{1}{1-k}$
14	$W(x)$	$2k-x$	$1-\frac{1}{2k-x}$	$\frac{1}{1-2k+x}$

В первых строках показано, как при смене окружностей меняются координаты выбранной точки (произвольной точки x и точек касания окружностей A, B, C). В восьмой строке парой точек k_1 и k_2 задаётся

некоторая окружность. Далее вычисляется, как выглядит симметрия относительно пары точек (т. е. относительно проходящей через них окружности) в разных случаях. Обратим внимание на строку 9. Заполнить две недостающие её графы я предлагаю читателю самостоятельно. Эта формула ранее не приводилась. Её несложно получить, например инвертируя I в окружность, а можно использовать свойства функции g. Эта формула требует осмысления. Во-первых, обратим внимание, что $H(x)$ — частное линейных функций, т. е. дробно-линейное преобразование. Во-вторых, $H(x)$ симметрично относительно k_1 и k_2. Это естественно, ведь окружность H задаётся парой точек k_1 и k_2 и не важно, в каком порядке они указаны.

В-третьих: введём обозначение

$$y = \frac{2k_1k_2 - x(k_1 + k_2)}{k_1 + k_2 - 2x}.$$

Здесь y — координата точки, симметричной x относительно H. Избавимся от знаменателя, перенесём всё в правую часть, получим

$$0 = 2xy - (x + y)(k_1 + k_2) + 2k_1k_2.$$

Полученное выражение симметрично не только относительно k_1 и k_2, но и относительно x и y. Это неудивительно, ведь $H(x)$ — симметрия, поэтому в уравнение, связывающее $H(x) = y$ и x, переменные x и y *должны* входить симметрично. Ранее было указано, что всякая композиция симметрий представима дробно-линейным преобразованием. Пользуясь этим рассуждением (но начав с другого конца, с общей симметричной зависимости x и y), легко показать, какие дробно-линейные преобразования задают симметрию.

В каких случаях $x = y$ (т. е. $H(x) = x$)? Подставив $x = y$ в уравнение, получаем: $0 = x^2 - x(k_1 + k_2) + k_1k_2$. Решая квадратное уравнение, получаем два корня $x = k_1$, $x = k_2$. Результат легко было предсказать: точки k_1 и k_2 в самом деле неподвижны при симметрии относительно H. Собственно, именно таким образом и была найдена формула строки 9. Теперь обратим внимание на следующее: мы начали рассуждение с точек k_1 и k_2. Поэтому полученное уравнение заведомо имеет решение в вещественных числах. Но мы могли бы начать с дробно-линейного преобразования или с какого-то симметричного уравнения, связывающего x и y. Тогда, находя неподвижные точки этого преобразования, мы бы получили квадратное уравнение $0 = x^2 - vx + w$. Если это уравнение имеет решения в действительных числах, то они дают координаты неподвижных точек. А если решения — комплексные числа? Есть ли какой-то геометрический смысл у этих комплексных решений? Оказывается — есть. Решения можно изобразить парой

точек, симметричных относительно I, а само преобразование оказывается биплетной симметрией относительно этой пары точек (точки на I при такой симметрии переходят друг в друга).

Мы пришли к двум важным темам: биплетная симметрия и связь эстетической геометрии с комплексными числами. Тема биплетной симметрии подробней рассмотрена на сайте [7]. В статье же даётся определение симметрии относительно пары точек или биплета только для частного случая, когда все точки лежат на I. Это же определение можно дать на основе гармонического разделения точек. В исчислении биплетов (то есть в исчислении композиции симметрий относительно них) важнейшую роль играет понятие перпендикулярных или коммутирующих биплетов. По сути это переименование гармонического разделения: два биплета перпендикулярны (коммутируют), если точки одного из них симметричны относительно другого. Тут действует *принцип взаимности*. Зачем же переименовывать симметричные или гармонические биплеты в «перпендикулярные»? Дело в том, что свойства симметрий биплетов похожи на свойства симметрий относительно прямых в трёхмерном евклидовом пространстве. И полностью совпадают со свойствами симметрий относительно прямых в трёхмерном пространстве Лобачевского.

В трёхмерном пространстве для двух прямых в общем случае существует одна прямая, перпендикулярная им обеим. Точно так же для двух биплетов существует третий, перпендикулярный (коммутирующий) с ними обоими. (За одним исключением — когда у пар точек, образующих два биплета, есть общая точка: тогда нет симметрии, коммутирующей одновременно с биплетами (v, w) и (v, z)). Заметим, что композиция симметрий относительно двух перпендикулярных биплетов сама является биплетной симметрией (так же как композиция симметрий относительно двух перпендикулярных прямых есть симметрия относительно перпендикулярной им обеим прямой), а шесть точек, образующих три упомянутых биплета, являются «великолепной шестёркой» точек, разбираемой подробно на сайте [7]: это точки касания четырёх взаимокасающихся окружностей или точки пересечения трёх взаимоперпендикулярных. Последним свойствам биплетной симметрии уже нет аналогов в мире прямых.

Наличие у двух биплетов перпендикулярного им позволяет сразу установить важнейшее свойство функции g. Вернёмся к рис. 13, где мы сравниваем изгибы на четырёх точках p, s, q, x. Утверждалось, что значение функции g не меняется, если поменять местами две пары переменных. Сейчас мы более геометрично объясним, почему это происходит. Пусть мы меняем местами p и s, q и x. Существует биплет, меняющий местами указан-

ные точки, — это биплет, перпендикулярный биплетам (p, s) и (q, x). Но при биплетной симметрии все изгибы сохраняются, все свойства симметричных точек совпадают. Следовательно, значение функции g не изменится.

Рассмотрим (на плоскости или на окружности I) совокупность биплетов, перпендикулярных данному. Оказывается, композиция симметрий относительно любых трёх из них — снова биплетная симметрия. Рассмотрим совокупности композиций относительно любых пар биплетов, коммутирующих с данным. Все полученные композиции коммутируют друг с другом. Они образуют группу, которая изоморфна группе умножения комплексных чисел (если мы рассматриваем только биплеты, лежащие на I, то будет изоморфизм с группой умножения действительных чисел). Чтобы определить сложение, надо рассмотреть совокупность биплетов, у которых есть одна общая точка. Кстати, привычную центральную симметрию можно рассматривать как биплетную: это симметрия относительно пары точек, одна из которых перед нами (центр симметрии), а вторая бесконечно удалена. В биплетном исчислении много интересных возможностей. Заметим, что некоторые свойства плоской геометрии теряются при переходе в трёхмерное пространство: тогда совокупность биплетов, коммутирующих с данным, не обладает указанными выше свойствами. Но на основании этой совокупности также можно построить интересную структуру, видимо совпадающую с кватернионами.

Прежде чем отойти от «таблицы координат», укажем на её важность для плоской геометрии Лобачевского, а именно модели Кэли — Клейна. В ней каждой прямой геометрии Лобачевского ставится в соответствие хорда окружности I. У нас эта хорда определяется парой точек на I (определяющих биплетную симметрию или симметрию относительно окружности, перпендикулярной I). Таким образом, свойства симметрий относительно прямых геометрии Лобачевского сводятся к свойствам симметрий относительно пары точек, которые описываются строкой 9 таблицы.

Теперь подойдём к комплексным числам и геометрии окружности с другой стороны. Известно, что с помощью композиции симметрий любые три точки плоскости можно отобразить в любые три другие точки. Кстати, простейшее доказательство основано на рассмотрении окружностей, касающихся друг друга в этих точках. Оно дано на сайте [4] (статья 9, теорема об отображении трёх точек). Из этого следует, что с точки зрения геометрии окружности любые три точки равноправны. Если мы можем в рамках геометрии окружности сделать какое-то построение, используя три *данные* точки, то точно такое же построение даст точно такой же результат на *любых* трёх точках. А вот четвёрки точек отличаются друг

от друга в геометрии окружности. Мы видели, что для четырёх точек, лежащих на одной окружности, можно ввести «координаты» или изгиб. Каждой четвёрке точек на окружности функция g сопоставляет *вещественное* число, и это число является инвариантом данной четвёрки точек. После симметрий данная четвёрка точек перейдёт в какую-то другую четвёрку, но значение функции g не изменится. Соответственно, если $g(A, B, C, D) \neq g(A', B', C', D')$, то и невозможно композицией каких-либо симметрий отобразить одновременно точки A в A', B в B', C в C', D в D'. Если же точки A, B, C, D *не лежат* на одной окружности, то им можно сопоставить *комплексное* число, причём это сопоставление происходит при помощи функции g', по всем своим свойствам аналогичной функции g, только она определяется на комплексных числах и её значение — комплексное число. При этом значение функции g' для случая, когда четыре её аргумента лежат на одной окружности, совпадает со значением функции g.

Связать комплексные числа и эстетическую геометрию можно также с помощью двухцентровых спиралей. Такие спирали встречаются, например, в барокко. Их примеры и свойства приводятся на сайте [7]. Совокупность спиралей, имеющих одни и те же центры (неподвижные точки, откуда всё вытягивается и куда всё втягивается), можно рассматривать как множество комплексных чисел. Умножению комплексных чисел соответствует композиция преобразований, определяющих спирали, эти преобразования коммутируют между собой. А сложение комплексных чисел определяется через биплетную симметрию.

Связь с комплексными числами и кватернионами — интересный, но внешний для эстетической геометрии вопрос. Важней изучать эстетическую геометрию, исходя из её внутренней логики, отвечать на вопросы, возникающие внутри неё самой. Попутно это может прояснить связи между уже известными областями алгебры и геометрии. Рассмотрим следующий шаг в изучении изгиба и систем координат. Мы начали с трёх касающихся окружностей A, B, C. Вернёмся к ним, рассмотрим точку x на плоскости, у неё есть три координаты $(A–B)(x)$, $(B–C)(x)$, $(C–A)(x)$. Как было сказано, никакая из них не определяет однозначно две другие. Именно чтобы избавиться от неоднозначности, рассматривались не все точки плоскости, а только лежащие на окружности I. В этом случае по одной координате точки x можно однозначно восстановить две другие, что и показано в таблице. Но в общем случае надо знать *две* координаты, например $(A–B)(x)$ и $(B–C)(x)$, чтобы найти третью, $(C–A)(x)$ в данном случае. При этом, даже найдя третью координату (она уже находится однозначно), мы не находим однозначно саму точку x. У неё остаются *две* возможности

(поскольку в данном случае все три возникающие окружности проходят через две точки, по обобщённой теореме о трёх окружностях). Поэтому, чтобы точно указать точку x, надо воспользоваться ещё чем-то. Проведём четвёртую окружность D, касающуюся трёх исходных, и рассмотрим, например, $(A–D)(x)$. Зная и эту координату точки x, мы находим саму точку уже однозначно.

До этого мы шли от точки к координатам. Поставим противоположную задачу. Пусть мы знаем координаты и хотим по ним восстановить точку x. Это можно сделать не всегда. Пусть нам известна $(A–B)(x)$, это определяет окружность из пучка (A, B), на которой лежит x. Пусть нам известна ещё и $(B–C)(x)$, это определяет окружность из пучка (B, C), в которой лежит x. Но эти две окружности могут не иметь общих точек, и потому точки с такими координатами просто не существует! А что происходит, если эти окружности касаются? Точка x обязательно лежит на I, т. е. этот случай мы как раз рассмотрели. Если же окружности не имеют общих точек, то, хотя *точки* с такими координатами на плоскости нет, вопрос о том, какой же будет координата $(C–A)$, имеет смысл, поскольку существует единственная окружность из пучка (C, A), лежащая в том же *пучке*, что и две новопроведенные окружности. Мы видели многообразие возможных координат. Оно особенно интересно, если мы систематически включаем в рассмотрение окружность D (касающуюся трёх исходных) и рассматриваем пучки (A, D), (B, D), (C, D) и координаты-изгибы в них. Появляется задача: как описать все возникающие связи между координатами точки в этих пучках и что происходит, если координаты есть, а точки — нет (соответствующие окружности не имеют общих точек)? Рассматривать эту задачу стоит в свете теории групп, в терминах перестановок окружностей A, B, C, D между собой. Разумеется, задача обобщается на трёхмерный и многомерный случай. Там надо рассматривать изгибы сфер, касающихся друг друга, и в трёхмерном пространстве существует пять сокасательных сфер (в n-мерном пространстве таких сфер $n + 2$). Каждая пара сокасающихся сфер задаёт систему координат, в которой можно мерить изгибы сфер, лежащих в этом пучке. Какие есть связи между этими координатами в общем случае?

Этот вопрос имеет отношение к теории представлений. Обычное введение в задачи теории представлений требует больше математической техники и менее геометрично.

Всё проведённое исследование базировалось на тройственной симметрии, т. е. на том, что три исходные окружности (и любые их пары) обладают абсолютно одинаковыми свойствами. Аналогичное свойство есть и у че-

тырёх сокасающихся окружностей, и у сокасающихся сфер. Но именно аналогия, а не полное сходство имеет место. Скажем, точки касания трёх окружностей можно выбрать абсолютно произвольно. А уже точки касания четырёх окружностей (великолепная шестёрка) — вовсе не произвольны. Поэтому тройственная симметрия (а не пятерная или шестерная) занимает особое место.

Тройственная симметрия, правда иначе, проявляется и в геометрии Евклида. Общеизвестнейшие примеры — теоремы о пересечении высот, биссектрис, медиан. У всех этих теорем есть аналоги в геометрии окружности, из теорем о треугольнике они превращаются в теоремы о трёхокружнике, это разобрано на сайте [7]. Сейчас мы свяжем тройственную симметрию с биплетами, т. е. с симметрией относительно пары точек. Прежде всего заметим, что произвольная пара биплетов A и B задаёт пучок биплетов. Это совокупность таких биплетов X, что $A * B * X$ — снова инволютивное преобразование, симметрия относительно какого-то четвёртого биплета. Эту совокупность легко описать геометрически. Найдём биплет Y, перпендикулярный (коммутирующий с) A и B. Тогда *все* биплеты, перпендикулярные Y, лежат в пучке, созданном A и B. Мы уже делали подобное построение, говоря о связи геометрии окружности с комплексными числами. Теперь же мы с его помощью превратим три произвольных биплета в аналог треугольника, по крайней мере научимся строить высоты в «трехбиплетнике». Пусть даны биплеты A, B, C. Назовём высотой, опущенной на C, биплет c, лежащий в пучке (A, B) и перпендикулярный C. Нетрудно доказать, что такой биплет с существует и единственен. Аналогично определяются высоты на B и A. Теорема о перпендикулярах «трехбиплетника» утверждает, что три биплета-высоты a, b, c сами лежат в одном пучке («лежать в одном пучке» в данном случае аналогично понятию «пересекаться в одной точке» для прямых). В этой теореме особенно ярко выражается тройственная симметрия. Она имеет интересные обобщения на многомерное пространство: ведь три биплета могут не лежать в одной плоскости. Биплет состоит из двух точек, в общем случае в трёх биплетах 6 точек и они лежат в пятимерном пространстве. Что можно сказать о высотах этого пятимерного трехбиплетника? Вопрос заслуживает изучения. Как и обобщения тройственной симметрии высот и биссектрис.

Есть совершенно другой подход к геометрии окружности. Его можно назвать «абстрактно алгебраическим». В математике есть различные теории (например, теория полугрупп), где изучаются записи из букв конечного алфавита и возможные равенства между составленными таким образом словами. Геометрию окружности можно строить, начиная именно с после-

довательности слов трёх- или четырёхбуквенного алфавита. Далее нужно рассматривать симметрии слов, возникающие после замены во всех словах одной буквы на другую, как это показано в § 9. И включать в рассмотрение эти симметрии как новые буквы (или слова?). Получающуюся структуру можно геометризировать. Ключевыми здесь, вероятно, окажутся простые соображения: между композициями симметрий относительно окружностей возможны равенства только в особых геометрических ситуациях. Если есть всего две окружности, то эти равенства могут выполняться, только если окружности пересекаются. Если они касаются — есть равенство в пределе (при композиции, проведённой бесконечное число раз). Но пока можно лишь наметить эту тему.

Возникает вопрос об аксиоматизации геометрии окружности. Его можно решить довольно просто в рамках классического подхода. Можно воспринимать геометрию окружности как раздел геометрии Лобачевского. На сайте [7] показано, как исходя из геометрии окружности определить неевклидовы геометрии. Чаще поступают наоборот, хотя этот путь и более сложен и менее нагляден. Плоскую геометрию окружности можно определить через трёхмерную геометрию Лобачевского. Но мне этот подход представляется неудачным: аксиомы должны как-то пояснять суть предмета, а при таком методе аксиоматизации ничего подобного не происходит. Мне представляется вопрос об аксиоматизации преждевременным, интереснее исследовать геометрию окружности как таковую и в связи с другими разделами математики. А может быть, подходящих со всех точек зрения аксиом и не выбрать: в конце концов у самой же окружности нет начала.

Я наметил несколько интересных вопросов, связанных с эстетической геометрией и её местом в математике. Укажу на её возможные связи с другими науками. Многие, видевшие образы, сделанные методами эстетической геометрии, «бесконечные картины» видеоарта программы dodecaLook высказывались о её значении для биологии. В самом деле, антропоморфные и зооморфные образы возникают в геометрии окружности, когда их и не ждёшь. Я полагаю, что законы эстетической геометрии пригодятся для понимания морфогенеза. С другой стороны, приведённая в §§ 5, 9, 10 связь геометрии окружности с простейшими знаковыми системами наводит на мысль о её полезности для семиотики. Я не специалист в обеих науках и не берусь подробно развивать эти соображения, но хочу привлечь к ним внимание.

И, разумеется, есть возможности для применения эстетической геометрии к классической эстетике и к психологии восприятия нами красоты.

Есть огромный простор для поиска закономерностей в произведениях искусства: живопись, скульптура, архитектура в свете симметрии относительно окружности. Особенно напрашиваются связи образов эстетической геометрии с архитектурой барокко и религиозным искусством Востока, например мандалами. Я полагаю, что систематическое изучение изобразительных искусств подобными методами может далеко расширить наше знание о красоте, а это — интересная и достойная задача для науки.

Список литературы

[1] *Бахман Ф.* Построение геометрии на основе понятия симметрии. М.: Наука, 1969.

[2] *Мамфорд Д., Райт Д., Сирис К.* Ожерелье Индры. Видение Феликса Клейна. М.: МЦНМО, 2011.

[3] *Пименов Р. Р.* Эстетическая геометрия, или Теория симметрий. СПб.: Школьная лига, 2014.

[4] http://bogemnyipeterburg.net/revolt/matem/Bereg_site/index.html

[5] http://bogemnyipeterburg.net/revolt/matem/index.htm

[6] http://bogemnyipeterburg.net/revolt/matem/pict/index.html

[7] http://bogemnyipeterburg.net/revolt/matem/teachpictures/index.html

[8] http://bogemnyipeterburg.net/revolt/matem/Bereg_site/bereg_4.html

[9] http://www.centre.smr.ru/win/pics/pic0084/p0084_800.htm

Р. Р. Пименов, ФТШ им. Иоффе «Академический Университет», Санкт-Петербург

revoltp@mail.ru

Повесть о трёх кругах

Г. Гальперин, Ч. Делман

Все знают, что сумма углов треугольника, образованного тремя прямыми на плоскости, равна 180°. А что можно сказать о сумме углов криволинейного треугольника, образованного тремя окружностями? Приглашаем читателя на опыте убедиться, что ответы могут быть самыми разными, а ситуация в целом неочевидна. В этой статье мы дадим полный ответ на поставленный вопрос. Как мы надеемся, в результате нашего исследования нам удастся раскрыть проблему более глубоко — рассмотрев её с нескольких точек зрения и на нескольких уровнях абстракции.

Мы начнём с элементарного решения, использующего лишь начальные понятия евклидовой геометрии. Прямое и очень короткое, это решение, однако, неполно. Оно годится лишь для одного специального расположения окружностей на плоскости. В модели гиперболической геометрии (плоскости Лобачевского) мы найдём ключ к другому частному случаю, что наводит на мысль, что различные реализации этой задачи лежат в континууме моделей геометрий с меняющейся кривизной. Этот более широкий геометрический подход приводит ко многим красивым обобщающим идеям и полностью решает исходную задачу единым методом. В конце статьи мы приводим ещё одно очень простое решение, основанное на преобразованиях плоскости, — оно, по нашему мнению, очень удачно завершает затронутую тему, поскольку каждое такое преобразование плоскости можно рассматривать просто как изменение точки зрения на рассматриваемый предмет. Предшествующие рассмотрения углубляют наше понимание открывшейся перспективы и позволяют описать нужные нам преобразования чисто геометрически.

Для читателя, незнакомого с классическими неевклидовыми геометриями, в которых по-новому истолковываются понятия *прямая* и *расстояние*, мы приводим соответствующий обзор, почти полностью замкнутый логически. Читателю, с этими геометриями знакомому, предлагается посмотреть на них по-новому: мы уверены, что его ждут неожиданные открытия.

Три задачи

Рассмотрим три окружности на плоскости, пересекающиеся трансверсально (т. е. без касания) в одной и той же точке P (рис. 1). Чему равна сумма углов треугольника ABC, образованного дугами окружностей? Ответ: 180°! (Почему? — постарайтесь понять это, просто глядя на картинку.)

Теперь рассмотрим три окружности с коллинеарными центрами (рис. 2). Получилось два (очевидно, конгруэнтных) криволинейных треугольника. Что можно сказать о сумме углов в этом случае? Ответ: на этот раз сумма меньше 180°!

Наконец, рассмотрим три окружности, пересекающиеся как на диаграмме Венна общего вида (рис. 3). Граница пересечения всех трёх кругов — выпуклый криволинейный треугольник; сумма его углов больше 180°, поскольку прямолинейный треугольник с теми же вершинами лежит внутри него. Что можно сказать об остальных шести криволинейных треугольниках, образованных этими окружностями? Ответ: как может догадаться читатель, хотя это совсем не очевидно из диаграммы, такая сумма всегда больше 180°!

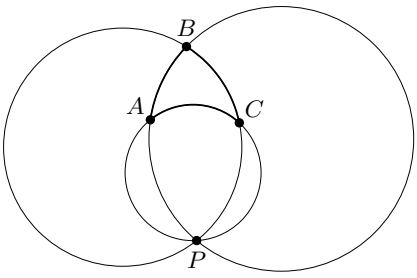


Рис. 1. Три окружности, проходящие через общую точку

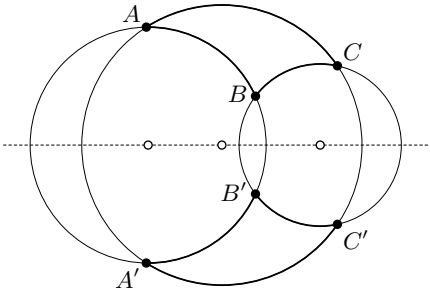


Рис. 2. Три окружности с коллинеарными центрами

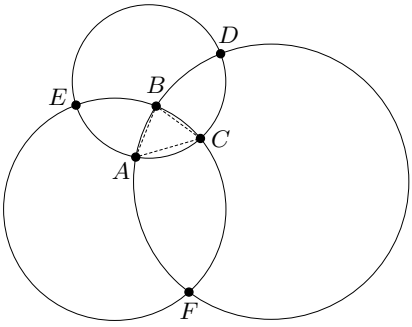


Рис. 3. Три окружности на диаграмме Венна

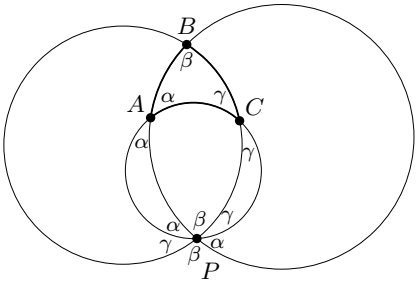


Рис. 4. Элементарное решение первой задачи

Первая из перечисленных задач допускает элементарное решение. Рассмотрим треугольник ABC на рис. 4. Заметим, что по соображениям симметрии все углы, обозначенные буквой α, равны между собой, и то же верно для углов, обозначенных через β и γ. Поэтому $2(\alpha+\beta+\gamma)=360^\circ$, откуда $\alpha+\beta+\gamma=180^\circ$.

Вторая задача также допускает простое решение, хотя и требующее более продвинутой геометрической идеи. А именно, полуплоскость, лежащая над линией трёх центров, может рассматриваться как верхняя полуплоскость гиперболической геометрии. Полуокружности с центрами на её границе — это (гиперболические) прямые, (гиперболические) углы между которыми совпадают с евклидовыми углами между полуокружностями; а сумма углов любого гиперболического треугольника меньше 180°.

В третьей задаче мы также покажем, что окружности являются *прямыми* в модели некоторой геометрии, на этот раз сферической, в которой сумма углов любого треугольника больше 180°. В общем случае мы рассмотрим произвольную конфигурацию трёх попарно пересекающихся окружностей и дадим простой критерий для выбора геометрии, которой эта конфигурация соответствует.

Основная теорема

Ключ к выбору геометрии, описывающей конфигурацию окружностей, — расположение точки пересечения прямых, соединяющих точки попарного пересечения окружностей. А именно, справедлива следующая лемма.

Лемма. *Пусть c_1, c_2 и c_3 — три окружности на плоскости, причём каждая пара окружностей пересекается в двух различных точках. Пусть l_{12}, l_{13} и l_{23} — прямые, заданные этими попарными пересечениями. Тогда прямые $\{l_{ij}\}$ либо конкурентны в точке P (т. е. все пересекаются в этой точке), либо параллельны (в последнем случае мы считаем их* конкурентными на бесконечности*). При этом есть ровно три возможности расположения точки конкурентности P по отношению к окружностям*:

1) *P лежит на всех трёх окружностях* (рис. 5 *а*);
2) *P лежит вне всех трёх окружностей* (рис. 5 *б*), *возможно на бесконечности*;
3) *P лежит внутри всех трёх окружностей* (рис. 5 *в*).

При этом если точка P лежит вне всех трёх окружностей, но не на бесконечности, то касательные из P к каждой окружности имеют равную длину. Окружность с центром P и радиусом этой общей длины перпендикулярна ко всем трём исходным окружностям. Если же P лежит

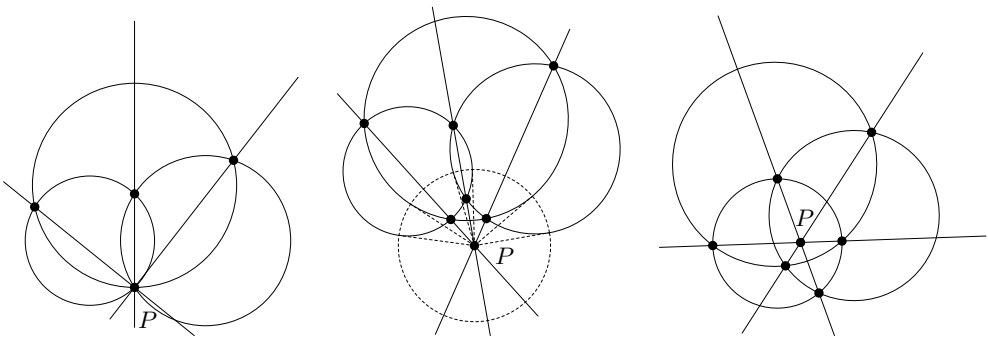


Рис. 5. Три возможности расположения точки P

на бесконечности, то эту роль играет прямая, проходящая через центры всех трёх окружностей (см. рис. 2).

Доказательство. Пусть прямая, проходящая через P, пересекает некоторую окружность c в точках A и B (не обязательно различных). *Степенью точки P относительно окружности c* называется взятое со знаком произведение $(PA) \cdot (PB)$, где (PA) обозначает направленное расстояние от P до A. Предлагаем читателю доказать, используя подобие треугольников, что степень точки не зависит от выбора прямой. (Доказательство можно найти в книге Коксетера и Грейтцера [3, теорема 2.11].)

Степень точки P относительно c положительна, если P лежит вне c, отрицательна, если P лежит внутри c, и равна 0, если P лежит на c. Рассматривая PA и PB как векторы, можно равносильно определить степень как скалярное произведение $PA \cdot PB$; поскольку векторы лежат на одной прямой, косинус угла между ними равен ± 1 в зависимости от того, одинаково их направление или противоположно.

Если две окружности пересекаются в точках A и B, мы получаем (рассматривая, например, прямую $\overleftrightarrow{PA}$), что степень точки P относительно обеих окружностей одинакова в точности тогда, когда точка лежит на прямой $\overleftrightarrow{AB}$. Пусть прямые l_{12} и l_{23} не параллельны, и пусть P — точка их пересечения. Тогда степень точки P относительно всех трёх окружностей одинакова; следовательно, P лежит на l_{13} и все три прямые конкурентны. Если же, с другой стороны, две прямые параллельны, то параллельны и все три, что доказывается от противного.

Если прямые пересекаются, тогда, как мы уже отметили, степень точки P относительно всех трёх окружностей одинакова. Обозначая эту общую степень через $\mathcal{P}$, получаем:

Случай 1. Если $\mathcal{P} = 0$, то P лежит на всех трёх окружностях.
Случай 2. Если $\mathcal{P} > 0$, то P лежит вне всех трёх окружностей.
Случай 3. Если $\mathcal{P} < 0$, то P лежит внутри всех трёх окружностей.

При этом если P лежит вне окружностей, то длина касательной из P к любой из них равна $\sqrt{\mathcal{P}}$. Окружность с центром P и радиусом $\sqrt{\mathcal{P}}$ ортогональна всем трём исходным окружностям. □

Отклоняясь от нашей темы, интересно отметить, что для произвольной пары окружностей, не обязательно пересекающихся, точка имеет одинаковую степень относительно обеих в точности тогда, когда она лежит на специальной прямой, которая называется их *радикальной осью* и ортогональна отрезку, соединяющему их центры. Выписывая уравнение этой прямой в декартовых координатах, Коксетер и Грейтцер [3, § 2.2] доказывают красивый факт: *если уравнение окружности приведено к стандартному виду*

$$F(x, y) = (x - a)^2 + (y - b)^2 - r^2 = 0,$$

то степень любой точки (x, y) *относительно этой окружности равна в точности* $F(x, y)$! Этот результат нельзя считать неожиданным, поскольку степень постоянна на окружностях, концентричных с данной.

Теперь мы готовы сформулировать основную теорему. Под *треугольником* мы будем понимать криволинейный треугольник, внутри которого не проходит дуга никакой из трёх данных окружностей. В случае 1 рассматриваются лишь треугольники, для которых P не является вершиной.

Теорема. *Пусть* c_1, c_2 *и* c_3 *— три окружности на плоскости, каждая пара которых пересекается в двух различных точках, а* l_{12}, l_{13} *и* l_{23} *— прямые, попарно соединяющие эти точки пересечения и конкурентные в точке* P. *Тогда сумма углов любого треугольника, образованного этими окружностями, удовлетворяет следующим условиям*:

1) *если* P *лежит на всех трёх окружностях, то сумма равна* 180°;
2) *если* P *лежит вне всех трёх окружностей, то сумма меньше* 180°;
3) *если* P *лежит внутри всех трёх окружностей, то сумма больше* 180°.

Для доказательства теоремы требуются некоторые сведения о классических неевклидовых геометриях, к которым мы теперь и обратимся.

Прогулка по трём геометриям

Геометрия — это абстрактная математическая теория, в которой неопределяемые понятия *точки* и *прямые* обладают определёнными свойствами. (Книга [5] даёт живое представление об историческом развитии классических геометрий. Другой весьма содержательный источник — книга [7].)

На практике мы делаем геометрию наглядной, работая с моделью, описанной более конкретно. В этой модели геометрические объекты представляют точки и прямые таким образом, что выполняются аксиомы рассматриваемой геометрии.

Декартова плоскость, состоящая из пар вещественных чисел, служит моделью евклидовой геометрии. Каждая пара чисел представляет точку, а прямые — это множества решений линейных уравнений. Декартова плоскость и её физические модели (например, поверхность стола), в которых расстояние между параллельными прямыми постоянно, а угол между ними и общей секущей одинаков, традиционно считаются *неискривлёнными*, в противовес сфере и другим *искривлённым* поверхностям. Поэтому мы говорим, что декартова плоскость имеет *нулевую кривизну*.

Сферическая геометрия. Любая поверхность, на которой введена функция расстояния, или *метрика*, является моделью некоторой геометрии. (Эти понятия обобщаются и на высшие размерности.) Прямые на такой поверхности, так называемые *геодезические*, — это кривые, наикратчайшие локально (т. е. для пар близко расположенных точек).

Сфера с унаследованной от трёхмерного евклидова пространства метрикой — это модель, в которой прямые — *большие окружности*, высекаемые плоскостями, проходящими через центр сферы. Слово *локально* в определении геодезической существенно; кратчайший путь выбирается между близкими точками.

Геодезичность больших кругов становится интуитивно ясной, если заметить, что длина дуги пропорциональна центральному углу, который она стягивает, а для центральных углов выполняется своего рода неравенство треугольника: когда пересекаются три плоскости, то сумма любых двух образовавшихся двугранных углов больше третьего. Поэтому путь, составленный из очень малых (можно считать, *бесконечно малых*) дуг, будет самым коротким, если все они лежат в одной плоскости, проходящей через центр.

Геодезичность больших кругов сферы можно объяснить и совсем наглядно, вовлекая в рассмотрение тот факт, что через любые две достаточно близкие точки на сфере проходит *единственная* геодезическая. Действительно, проведём через две произвольные близкие точки A и B большую окружность $\mathbf{c}$ и предположим, что на сфере найдётся более короткий, чем дуга $\smile AB$, путь γ с концами A и B. Отразив кривую γ относительно плоскости окружности $\mathbf{c}$, получим на сфере новую кривую γ' той же длины, что и γ. Поэтому из A в B будет вести не одна, а по крайней мере две геодезических — первая вдоль дуги γ, а вторая вдоль дуги γ', что противоречит единственности геодезической на сфере. Следовательно,

кривые γ и γ' должны совпадать и, тем самым, лежать не только на сфере, но и в плоскости окружности **c**, а значит, и на самой окружности **c**, сливаясь с дугой $\smile AB$.

Совсем же строгое доказательство, основанное на предыдущих рассуждениях, требует интегрирования и других элементов анализа. (Общедоступным введением в применение анализа к геометрии может служить книга [6].)

Исключительную важность для нас имеет сумма углов сферического (геодезического) треугольника. Углы удобнее вычислять, измеряя их в радианах, что мы в дальнейшем и будем делать. Приглашаем читателя убедиться экспериментально, что сумма углов сферического треугольника всегда превышает π и убывает вместе с площадью треугольника, приближаясь, но не достигая π, когда эта площадь стремится к нулю (и приближаясь, но не достигая 5π, когда треугольник заполняет почти всю сферу). Это наблюдение подсказывает, что надо сосредоточить внимание на разности между суммой углов сферического треугольника и π, которая носит название *углового эксцесса*. Докажем положительность эксцесса треугольника, подробнее исследовав его связь с площадью.

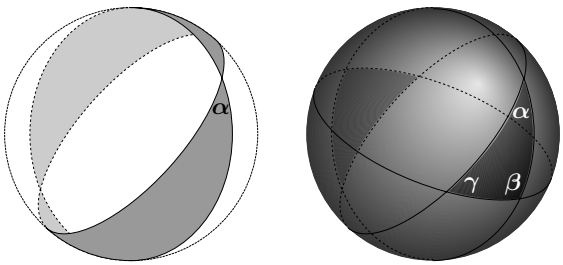
Нам при этом помогут следующие факты. Во-первых, заметим, что угловой эксцесс *аддитивен*: если разбить треугольник на два меньших, то, как нетрудно проверить, эксцессы частей в сумме дают эксцесс исходного треугольника. Во-вторых, *конгруэнтные треугольники заведомо имеют одинаковый эксцесс*. Но эти два факта — важнейшие свойства площади: площади частей некоторой фигуры в сумме дают площадь всей фигуры, а конгруэнтные фигуры имеют равную площадь!

Далее, сфера, подобно плоскости, *однородна*: любой треугольник можно переместить посредством вращений и отражений сферы в любую другую позицию как твёрдое тело, т. е. без изменения расстояний и углов (а также площади). В совокупности все эти факты подсказывают, что угловой эксцесс равен площади, умноженной на постоянный коэффициент.

Чтобы доказать, что угловой эксцесс пропорционален площади, заметим, что угол между двумя большими окружностями пропорционален площади сектора между ними. (Здесь существенна однородность сферы!) На сфере радиуса R площадь сектора с углом α равна

$$\frac{\alpha}{\pi}(4\pi R^2) = 4R^2\alpha$$

(рис. 6 *а*; для простоты мы одинаково обозначаем угол и его величину). Пусть Δ обозначает площадь треугольника с углами α, β и γ. Секторы с углами α, β и γ покрывают сферу с избытком; исходный треугольник и его антипод покрыты по три раза, а остальная часть сферы — ровно



а *б*

Рис. 6. Сектор, отсечённый сферическим углом, и угловой эксцесс треугольника

один раз (рис. 6 *б*) Отсюда несложными выкладками получаем формулу

$$\alpha + \beta + \gamma - \pi = \frac{\Delta}{R^2}.$$

В частности, на сфере радиуса $R = 1$ угловой эксцесс треугольника в точности равен его площади.

Гиперболическая геометрия. Сфера является поверхностью *постоянной положительной кривизны*. Существуют также поверхности постоянной *отрицательной* кривизны. Чтобы понять, как их можно описать, рассмотрим уравнение $x^2 + y^2 + z^2 = R^2$. Когда константа R^2 возрастает до $+\infty$, кривизна сферы с таким уравнением уменьшается до нуля. Можно считать евклидову плоскость сферой бесконечного радиуса. Что произойдёт, когда константа R^2 появится с другого, отрицательного конца числовой прямой?

Поверхность с уравнением вида $x^2 + y^2 + z^2 = -R^2 = (iR)^2$, которая формально является «сферой мнимого радиуса», далее называется *псевдосферой*. Разумеется, у этого уравнения нет решений, пока мы не расширим область значений переменных с вещественных на комплексные числа, а «радиус» псевдосферы получит смысл, лишь если обобщить понятие расстояния так, чтобы оно допускало мнимые значения.

Расстояния и углы в евклидовом пространстве $\mathbb{R}^n$ измеряются посредством *скалярного произведения*. Это пример симметричной билинейной формы — понятия, важного для многих областей математики. Кроме того, скалярное произведение положительно определено: для любого вектора $\mathbf{v}$ справедливо неравенство $\mathbf{v} \cdot \mathbf{v} \geqslant 0$, причём $\mathbf{v} \cdot \mathbf{v} = 0$ в точности тогда, когда $\mathbf{v} = 0$. Длина вектора $\mathbf{v}$ обозначается $|\mathbf{v}|$ и равна $\sqrt{\mathbf{v} \cdot \mathbf{v}}$, а угол θ между двумя векторами $\mathbf{v}$, $\mathbf{w}$ выражается формулой:

$$\cos\theta = \frac{\mathbf{v} \cdot \mathbf{w}}{|\mathbf{v}| \cdot |\mathbf{w}|}.$$

Если в качестве координат векторов допустить любые комплексные числа, скалярное произведение остаётся симметричной билинейной формой, но уже не положительно определённой. Форма с такими свойствами называется *псевдометрикой*. Эту форму мы и выберем для измерения расстояний и углов в комплексном пространстве $\mathbb{C}^n$, позволив этим величинам принимать комплексные значения.

Возвращаясь к уравнению $x^2 + y^2 + z^2 = -R^2$ и желая найти соответствующую двумерную поверхность, рассмотрим лишь те решения (x, y, z), в которых x и y вещественны. Оставшаяся переменная z тогда обязана быть чисто мнимой, так что положим $z = it$. Чтобы изобразить псевдосферу, начертим её образ в пространстве $\mathbb{R}^3$ при отображении $(x, y, it) \mapsto (x, y, t)$. Этот образ — двуполостный гиперболоид, множество решений уравнения $x^2 + y^2 - t^2 = -R^2$. Его компоненты аналогичны двум полусферам обычной сферы — северной и южной, — где начало координат $(0, 0, 0)$ можно считать центром, а точки $(0, 0, \pm R)$ — северным и южным полюсами сферы.

Чтобы измерять расстояния и углы, нужно помнить, что рассматриваемое пространство — образ подпространства из $\mathbb{C}^3$, где расстояния и углы измеряются посредством стандартного скалярного произведения. Поскольку вектор (x, y, t) в действительности представляет вектор (x, y, it), его настоящая длина равна $\sqrt{x^2 + y^2 + (it)^2} = \sqrt{x^2 + y^2 - t^2}$. Аналогично угол между двумя векторами (x_1, y_1, t_1) и (x_2, y_2, t_2) должен определяться формой $x_1x_2 + y_1y_2 - t_1t_2$. Это метрика Лоренца, трёхмерный вариант псевдометрики релятивистского пространства-времени. (Доступное и опирающееся на физику, но достаточно продвинутое введение в псевдометрики см. в книге [8].) Расстояния и углы, видимые на получившейся картине, отличаются от их истинных значений; например, истинная длина любого радиус-вектора, идущего от начала координат к поверхности, равна iR, т. е. мнимому числу!

Чтобы измерять расстояния и углы на самой псевдосфере, применим понятие псевдометрики к её *касательным векторам*. Для каждой точки P в пространстве $\mathbb{R}^3$ *касательное пространство в этой точке* — это экземпляр пространства $\mathbb{R}^3$, все векторы которого исходят из P. Касательная плоскость к поверхности в этой точке — двумерное подпространство этого пространства; прямая, касательная к поверхности в точке P, — одномерное подпространство касательной плоскости.

Полезна (и соответствует интуиции) запись координат и других величин, связанных с касательными векторами, в виде *дифференциальных* выражений. Если f — вещественнозначная функция, то её *дифференциал* — это функция касательного вектора, которая в каждой точке P даёт наилучшее линейное приближение к изменению функции f при перемещении вдоль касательного вектора, исходящего из P.

В частности, декартовы координаты x, y и t в нашей модели можно рассматривать как функции от $P \in \mathbb{R}^3$; пусть dx, dy и dt соответственно обозначают значения дифференциалов этих функций. Тогда dx, dy и dt, соответственно, дают первую, вторую и третью координаты касательного вектора после параллельного переноса осей координат в точку P (см. рис. 7 *а*). Отметим, что в данном случае линейные приближения dx, dy и dt оказываются точными, поскольку ортогональная проекция на прямую — линейное отображение. Поэтому (лоренцеву) длину касательного вектора удобно записать как дифференциальное выражение $\sqrt{dx^2 + dy^2 - dt^2}$.

Замечательно, что для векторов, касательных к псевдосфере, это дифференциальное выражение *вещественно и положительно определено*. Поскольку длина пути получается интегрированием длин его касательных векторов, пути между точками псевдосферы имеют вещественную, причём положительную, длину, и поэтому можно говорить о геодезических, т. е. о путях локально минимальной длины.

Чтобы убедиться, что $dx^2 + dy^2 - dt^2 > 0$, полезно использовать цилиндрические координаты. В качестве ортонормированного базиса касательного пространства к $\mathbb{R}^3$ в точке с цилиндрическими координатами (r, θ, t) возьмём единичный радиус-вектор, единичный вертикальный вектор и ортогональный им единичный вектор; в этом базисе координаты касательного вектора равны $dr, r\,d\theta, dt$. (Множитель r в угловой координате возникает потому, что перемещение по короткому пути вдоль окружности радиуса r меняет центральный угол на длину этого пути, умноженную на $1/r$, см. рис. 7 *б*.) Выражение $dx^2 + dy^2$, т. е. квадрат длины горизонтальной компоненты вектора, равно $dr^2 + r^2\,d\theta^2$. Для вектора, касательного к гиперболоиду, имеем $dt^2 < dr^2$, поскольку радиальный наклон гиперболоида меньше, чем у его асимптотического конуса. (На алгебраическом языке: из уравнения $r^2 - t^2 = -R^2$ следует, что $r^2 < t^2$; дифференцируя, получаем $dt^2 = \frac{r^2}{t^2}\,dr^2$.) Следовательно,

$$dx^2 + dy^2 - dt^2 = r^2\,d\theta^2 + (dr^2 - dt^2) > 0.$$

Геодезическими при этом оказываются пересечения псевдосферы с плоскостями, проходящими через начало координат, как и на обычной сфере, хотя здесь в этом труднее убедиться, и мы не будем это доказывать. Псевдосфера тоже однородна, а сумма углов треугольника выражается вполне аналогичной формулой:

$$\alpha + \beta + \gamma - \pi = \frac{\Delta}{-R^2}.$$

Как следствие, эта сумма всегда меньше π. (В книге [2] подробно рассмотрена как сфера, так и псевдосфера, см. стр. 74–78.)

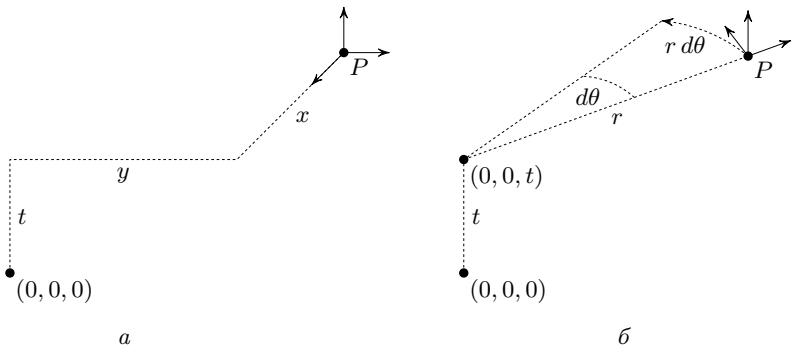


Рис. 7. Сравнение двух базисов касательного пространства в точке P

Каждая полусфера псевдосферы служит моделью геометрического объекта, который называется *гиперболической плоскостью*. В следующем разделе описано отображение псевдосферы в плоскость, которое сохраняет величины углов, давая нам полезную картинку. Аналогичное отображение существует и для обычной сферы. Отображение, сохраняющее углы, называется *конформным*.

Стереографическая проекция

Конформное отображение обычной сферы на плоскость получается проектированием из произвольной точки сферы на плоскость, касательную к ней в противоположной точке. Любое такое отображение, а также обратное отображение плоскости на сферу, называется *стереографической проекцией*. Каждая окружность на сфере (а не только большие окружности) стереографически проектируется на окружность или прямую (последнюю можно считать окружностью, проходящей через бесконечно удалённую точку ∞) в этой плоскости. При этом она отображается в прямую в точности тогда, когда проходит через центр проектирования (который отображается в ∞). Обратно, любая прямая или окружность на плоскости является образом окружности на сфере (рис. 8).

Важное замечание. Центр окружности на плоскости в общем случае является не проекцией центра соответствующей окружности на сфере, а проекцией *вершины конуса*, касательного к сфере в точках этой окружности.

Указанные важные свойства стереографической проекции могут быть доказаны с использованием элементарной геометрии. (Соответствующие рассуждения проведены и удачно проиллюстрированы в классической книге [1, стр. 248–251].)

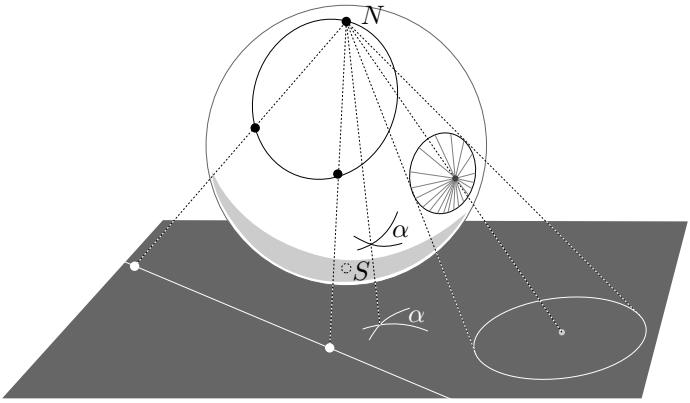


Рис. 8. Стереографическая проекция сохраняет углы и переводит окружности в окружности или прямые

Чтобы убедиться, что стереографическая проекция конформна, рассмотрим рис. 9 *а*, где показано сечение сферы плоскостью, проходящей через центр проектирования N, противоположную ему точку S и ещё одну точку P на сфере; $\pi(P)$ — проекция точки P, а A — точка, в которой сечение плоскости, касательной к сфере в точке P, пересекает плоскость, касательную в точке S. Соответствующая трёхмерная конфигурация изображена на рис. 10. Рассмотрим угол с вершиной P, касательный к сфере (он обозначен стрелками). Поскольку $\angle\pi(P)PA = \angle P\pi(P)A$, образ этого угла при проекции — его отражение относительно плоскости Π, проходящей через A и перпендикулярной к прямой $\overleftrightarrow{NP\pi(P)}$. Ясно, что эти углы равны.

На рис. 9 *б* показано локальное поведение стереографической проекции. Если Q — близлежащая точка на касательной плоскости через P, то $\triangle NQP$

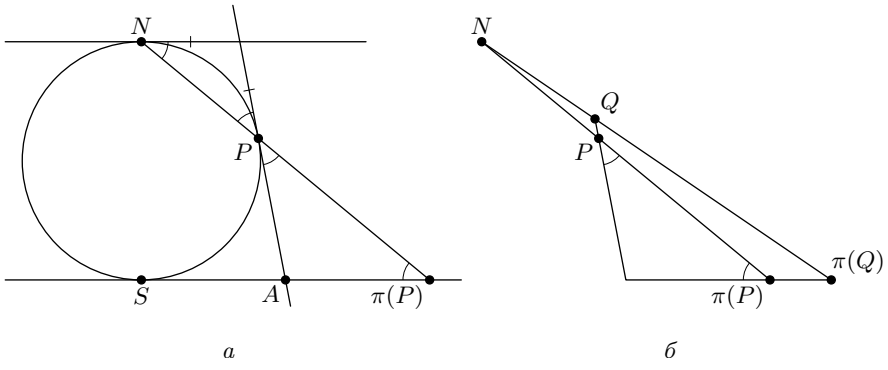


Рис. 9. Стереографическая проекция конформна и локально аппроксимируется растяжением

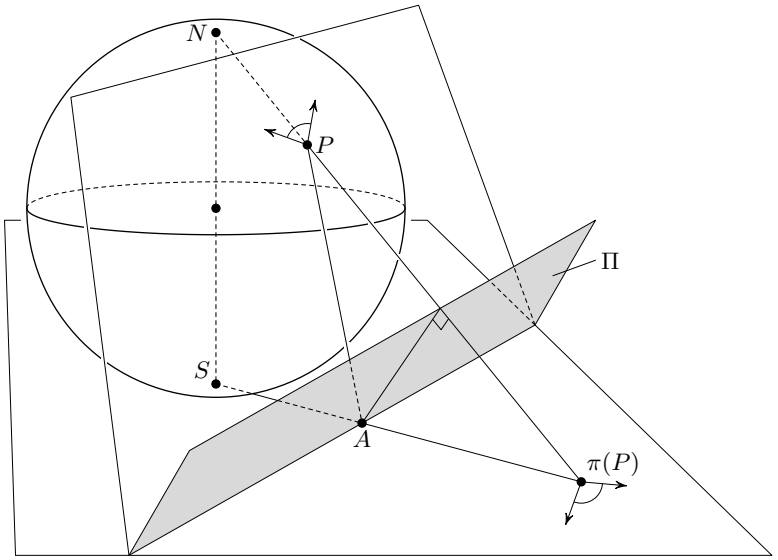


Рис. 10. Трёхмерное изображение: плоскости, касательные в точках P и S, являются отражениями друг друга относительно плоскости П, *перпендикулярной прямой, проходящей через точки N, P и $\pi(P)$*

приблизительно подобен $\triangle N\pi(P)\pi(Q)$. Следовательно, стереографическая проекция линейно аппроксимируется в каждой точке P гомотетией (растяжением). Коэффициент гомотетии меняется с широтой точки P, неограниченно возрастая при приближении P к N и достигая минимального значения 1 в точке S.

Чтобы получить теперь стереографическую проекцию псевдосферы на плоскость, надо спроектировать каждую точку гиперболоида на горизонтальную плоскость, проходящую через южный полюс $(0, 0, -R)$, т. е. на плоскость $t = -R$, вдоль прямой, соединяющей его с северным полюсом $(0, 0, R)$ (рис. 11). Эта проекция отображает южную полусферу на внутренность круга радиуса $2R$ с центром в начале координат, а северная полусфера, за исключением северного полюса, отображается на внешность круга. Внутренность круга реализует *модель Пуанкаре в круге* для гиперболической плоскости, а её граница называется *бесконечно удалённой окружностью*. Можно показать, что каждая геодезическая отображается на окружность или прямую, ортогональную бесконечно удалённой окружности (за вычетом точек пересечения); геодезические, проходящие через полюса, переходят в прямые.

Конформность проектирования гиперболоида нельзя доказать элементарными геометрическими рассуждениями, поскольку углы на гиперболо-

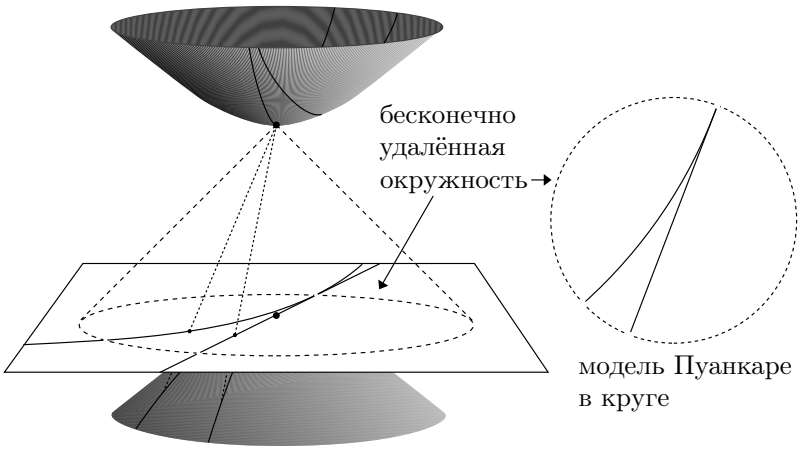


Рис. 11. Проекция на плоскость модели псевдосферы на гиперболоиде

иде измеряются иначе, чем на плоскости (с помощью другой билинейной формы). Поэтому, чтобы доказать сохранение величины углов, нужно прибегнуть к вычислениям и сравнить углы между касательными к кривым до и после проектирования. Читатель, склонный поверить нам и избежать технических подробностей, может пропустить последующую выкладку без всякой потери связности изложения. Для тех же, кто решится вникнуть в детали, приводимое ниже доказательство может послужить хорошим примером применения методов анализа в геометрии.

Доказательство конформности модели Пуанкаре в круге. Сосредоточим наше внимание на южной полусфере; доказательство для северной полусферы аналогично. Пусть π обозначает проектирование, распространённое на область $t < R$ и рассматриваемое как отображение на $\mathbb{R}^2$, без учёта последней координаты образа $(-R)$. Предположим, что кривые α и β пересекаются в точке $P = (x, y, t)$. Производная $D\pi(P)$ от π в точке P переводит векторы, касательные к α и β в этой точке, в векторы, касательные к их образам в точке $\pi(P)$ (согласно правилу дифференцирования сложной функции).

Пусть $\langle \mathbf{v}_P, \mathbf{w}_P \rangle_L$ — лоренцево произведение касательных векторов $\mathbf{v}_P$ и $\mathbf{w}_P$ в точке P. Докажем, что если $\mathbf{v}_P$ и $\mathbf{w}_P$ — касательные векторы к гиперболоиду, то

$$D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{w}_P) = \frac{4R^2}{(R-t)^2} \langle \mathbf{v}_P, \mathbf{w}_P \rangle_L.$$

Иными словами, скалярное произведение образов отличается лишь постоянным множителем от лоренцева произведения прообразов. Отсюда следует,

что угол θ между $\mathbf{v}_P$ и $\mathbf{w}_P$, задаваемый формулой

$$\cos\theta = \frac{\langle \mathbf{v}_P, \mathbf{w}_P \rangle_L}{\langle \mathbf{v}_P, \mathbf{v}_P \rangle_L^{1/2} \langle \mathbf{w}_P, \mathbf{w}_P \rangle_L^{1/2}},$$

равен углу ϕ между $D\pi(P)(\mathbf{v}_P)$ и $D\pi(P)(\mathbf{w}_P)$, для которого

$$\cos\phi = \frac{D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{w}_P)}{(D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{v}_P))^{1/2} (D\pi(P)(\mathbf{w}_P) \cdot D\pi(P)(\mathbf{w}_P))^{1/2}},$$

поскольку коэффициент пропорциональности в числителе и знаменателе этой дроби сокращается. В противоположность случаю сферы, здесь коэффициент пропорциональности *убывает* при удалении P от S (поскольку t убывает), принимая *максимальное* значение 1 в точке S.

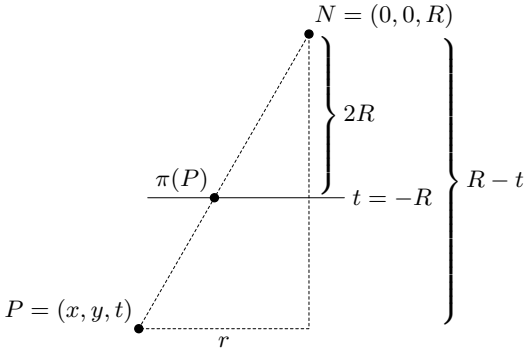


Рис. 12. Действие проектирования π на радиальную координату

Рассматривая подобные треугольники на рис. 12, мы видим, что образ точки P с цилиндрическими координатами (r, θ, t) при проектировании π — точка с полярными координатами $\left(\frac{2Rr}{R-t}, \theta\right)$. Производная в точке P от преобразования декартовых координат в цилиндрические переводит тройку $(dr, r\,d\theta, dt)$ в тройку $(dr, d\theta, dt)$. Если $\phi\colon \mathbb{R}^3 \to \mathbb{R}^2$ обозначает отображение $(r, \theta, t) \mapsto \left(\frac{2Rr}{R-t}, \theta\right)$, то его производная в точке (r, θ, t) есть

$$D\phi(r, \theta, t) = \begin{pmatrix} \frac{2R}{R-t} & 0 & \frac{2Rr}{(R-t)^2} \\ 0 & 1 & 0 \end{pmatrix}.$$

Применив $D\phi(r, \theta, t)$ к тройке $(dr, d\theta, dt)$ и вычислив затем производную в точке $\phi(r, \theta, t) = \left(\frac{2Rr}{R-t}, \theta\right)$ от обратного преобразования полярных координат в декартовы, получаем (опять же по правилу дифференцирования

сложной функции):

$$D\pi(P)(dr, r\,d\theta, dt) = \frac{2R}{R-t}\Big(dr + \frac{r\,dt}{R-t},\ r\,d\theta\Big).$$

Для любой точки P на гиперболоиде имеем $r^2 - t^2 = -R^2$, поэтому для любого вектора, касательного к гиперболоиду в точке P, выполняется равенство $dt = \frac{r}{t}dr$. После некоторых упрощений, использующих эти подстановки, получаем:

$$D\pi(P)\Big(dr, r\,d\theta, \frac{r}{t}dr\Big) = \frac{2R}{R-t}\Big(-\frac{R}{t}dr,\ r\,d\theta\Big).$$

Теперь читатель может проверить, что для двух касательных векторов в точке P выполняется искомое равенство:

$$D\pi(P)(\mathbf{v}_P) \cdot D\pi(P)(\mathbf{w}_P) = \frac{4R^2}{(R-t)^2}\langle \mathbf{v}_P, \mathbf{w}_P\rangle_L.$$

Последовательности стереографических проекций. Стереографическая проекция — незаменимое орудие преобразования геометрических моделей с сохранением углов между геодезическими. В частности, она даёт нам модель Пуанкаре гиперболической геометрии на полуплоскости.

Для этого сначала стереографически спроектируем круг Пуанкаре радиуса 2 на полусферу единичной сферы (а его дополнение, включая бесконечно удалённую точку, на другую полусферу) — с центром проекции в северном полюсе N. Затем стереографически спроектируем сферу на любую вертикальную плоскость, касающуюся экватора, — с расположенным на сфере в противоположной точке к точке касания центром проекции. Образ бесконечно удалённой окружности круга Пуанкаре при такой последовательности проекций называется *бесконечно удалённой прямой*. Каждая полуплоскость является образом гиперболической плоскости (рис. 13).

Применив две стереографических проекции из разных точек, мы получаем также конформное отображение евклидовой плоскости, в котором образы геодезических — прямые или окружности (рис. 14).

В итоге у нас теперь есть отображения плоскости, сферы и псевдосферы, при которых геодезические представлены прямыми и окружностями. Так же как географическую карту нужно исказить, чтобы напечатать её на плоских страницах атласа, наши отображения искажают действительные положения точек в изображаемых геометрических объектах. Искусство построения отображений вращается вокруг выбора проекции с подходящим типом искажения, который позволяет использовать отображение для намеченной цели. Например, известная меркаторская проекция земной

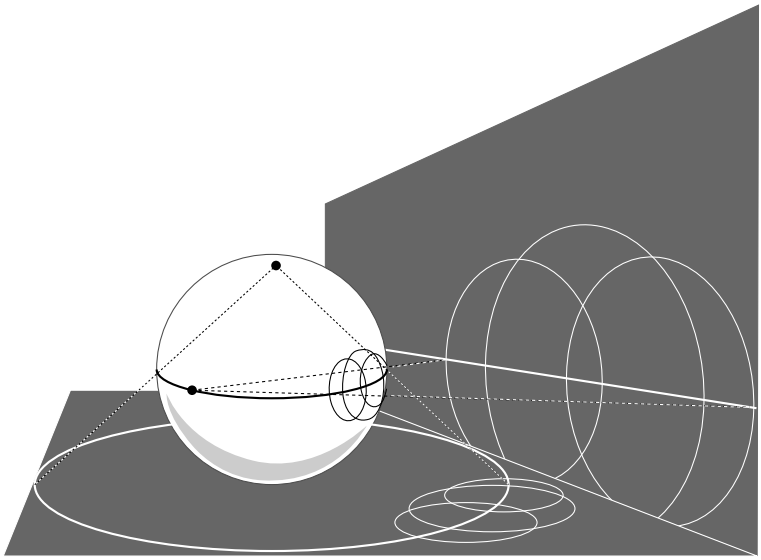


Рис. 13. Конформные преобразования между кругом Пуанкаре, полусферой и моделью Пуанкаре гиперболической геометрии на полуплоскости

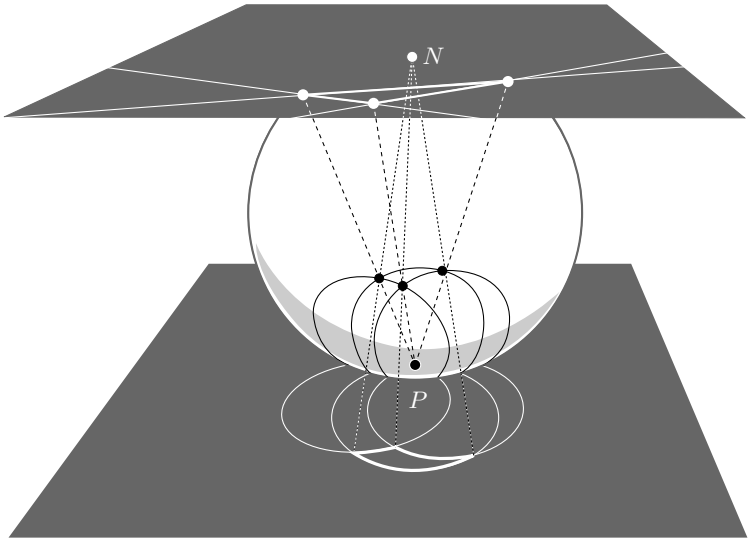


Рис. 14. Евклидова геометрия на сфере

поверхности полезна при навигации, поскольку она даёт правильный компасный пеленг между любыми двумя точками. (Мак-Клири [6, ch. 8^{bis}] даёт удачный обзор проекций для географических карт.) Поскольку нас интересует измерение углов, мы сосредоточились на конформных отобра-

жениях, сохраняющих углы между гладкими кривыми. Возникает вопрос: как узнать, имея тройку попарно пересекающихся окружностей, какое отображение мы видим? И вообще, могут ли все три окружности возникать в качестве геодезических в одной и той же модели?

Как узнать, какое отображение вы видите: доказательство теоремы

Случай 1: P лежит на всех трёх окружностях. В этом случае рассмотрим сферу, южным полюсом которой является точка P, и две плоскости, касательные к сфере в её южном и северном полюсах. Образы наших трёх окружностей на сфере — это окружности, проходящие через южный полюс. Если снова спроектировать, на этот раз из южного полюса сферы на вторую плоскость, то эти окружности перейдут в прямые линии, как на рис. 14. Таким образом, c_1, c_2 и c_3 — геодезические в конформной плоской модели евклидовой геометрии, и сумма углов образованного ими треугольника равна 180°.

Случай 2: P лежит вне всех трёх окружностей. В частном случае, когда прямые, соединяющие точки попарного пересечения окружностей, параллельны, их центры коллинеарны, что было доказано в первом разделе; поэтому будем считать, что существует точка конкурентности P этих прямых, лежащая вне всех трёх окружностей. В этом случае лемма обеспечивает существование окружности d, ортогональной всем трём окружностям. Тогда все дуги наших окружностей, лежащие целиком внутри или вне окружности d, — геодезические в круговой модели Пуанкаре гиперболической плоскости, и потому сумма углов любого образованного ими треугольника меньше 180°.

Случай 3: P лежит внутри всех трёх кругов. Рассмотрим семейство сфер (переменного размера), касающихся данной плоскости в точке P, которая является для них южным полюсом. Посмотрим, как площадь стереографической проекции круга с границей c_1 на каждую из этих сфер соотносится с площадью всей поверхности сферы: если сфера очень мала, проекция по площади больше половины сферы (действительно, она будет включать всю южную полусферу); если же сфера очень велика, проекция будет меньше половины всей поверхности. Всё зависит от того, лежит окружность c_1 внутри или же вне прообраза экватора на плоскости (рис. 15). Позволив сфере меняться непрерывно, мы видим, что существует единственная сфера $\mathcal{S}$, для которой площадь проекции в точности равна половине площади всей поверхности. (Другое рассуждение см. ниже в замечании 3.) Тогда образ окружности c_1 на сфере $\mathcal{S}$ — большая окружность.

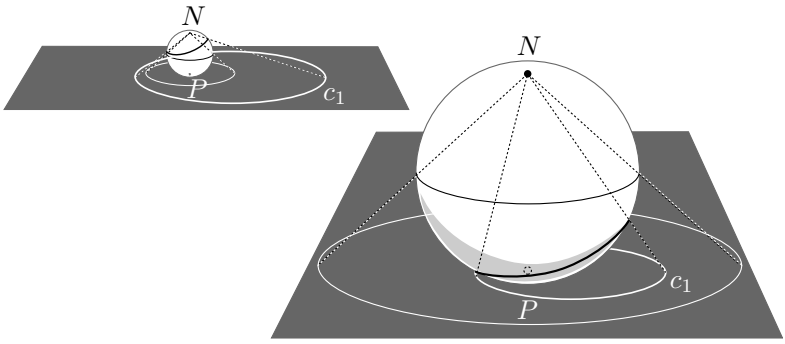


Рис. 15. Семейство сфер, касательных к плоскости в точке P

Мы утверждаем, что образы окружностей c_2 и c_3 на сфере $\mathcal{S}$ — тоже большие окружности. Достаточно доказать это для c_2, поскольку то же рассуждение годится и для c_3. Заметим, что образ прямой l_{12} при стереографической проекции — это *меридиан*, т. е. большая окружность, проходящая через северный и южный полюсы сферы. Поскольку образ окружности c_1 — тоже большая окружность, образы точек пересечения c_1 с l_{12} противоположны на сфере (антиподальны). А так как две точки пересечения $c_1 \cap l_{12}$ лежат также и на окружности c_2, образ этой окружности — большая окружность (рис. 16).

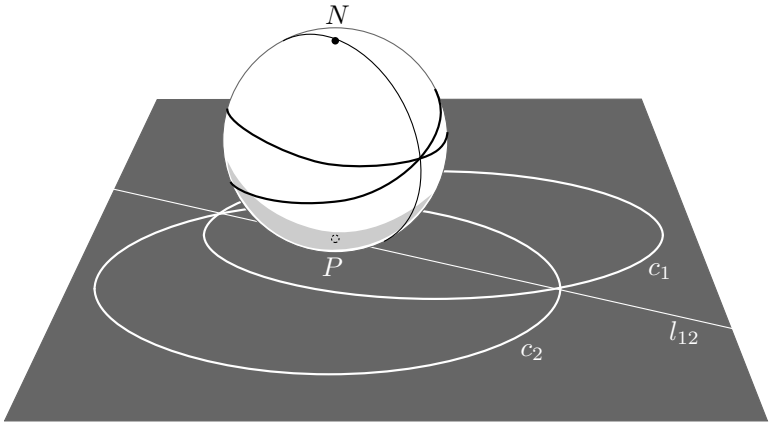


Рис. 16. Образы окружностей на сфере $\mathcal{S}$ — большие окружности

Итак, мы показали, что образы всех трёх окружностей на сфере $\mathcal{S}$ — большие окружности, т. е. геодезические. Таким образом, c_1, c_2 и c_3 — геодезические в конформной плоской модели сферической геометрии, следовательно, сумма углов любого образованного ими треугольника больше 180°. □

Замечание 1. Тот факт, что евклидова геометрия возникает лишь в случае, когда точка P лежит в точности на всех трёх окружностях, наглядно показывает, что наша привычная «плоская» геометрия — всего лишь одна точка в спектре геометрий (различной кривизны).

Замечание 2. Используя последовательность стереографических проекций, описанную в предыдущем разделе, можно преобразовать общую картину случая 2 в изображение на верхней полуплоскости.

Замечание 3. Хотя в случае 3 авторам нравится рассуждение с непрерывностью, от неё можно избавиться следующим образом. Пусть $s = \sqrt{-\mathcal{P}}$ (напомним, что $\mathcal{P}$ — степень точки P относительно каждой из трёх окружностей). Мы утверждаем тогда, что сфера $\mathcal{S}$ в проведённом выше доказательстве имеет радиус $s/2$. Чтобы показать это, рассмотрим для каждой из окружностей c_1, c_2 и c_3 проходящую через точку P хорду, перпендикулярную проходящему через P диаметру. Каждая такая хорда имеет длину $2s$, причём P — её середина. Проведём окружность e с центром P и радиусом s. Эта окружность при стереографической проекции переходит в экватор сферы $\mathcal{S}$, а окружности c_1, c_2 и c_3, каждая из которых пересекает окружность e в её диаметрально противоположных точках, обязаны, в силу этой причины, спроектироваться в большие окружности на сфере $\mathcal{S}$.

Это альтернативное рассуждение подчёркивает параллельность случаев 2 и 3.

Замечание 4. Теорема остаётся верной, если *окружность* понимается в широком смысле, т. е. допускается и прямая.

Приглашаем читателя обобщить теорему на предельные случаи, когда две или более окружностей взаимно касаются. Возможные конфигурации продемонстрированы на рис. 17. Случай I — евклидов. Одна вершина треугольника отправлена на бесконечность отображением, переводящим

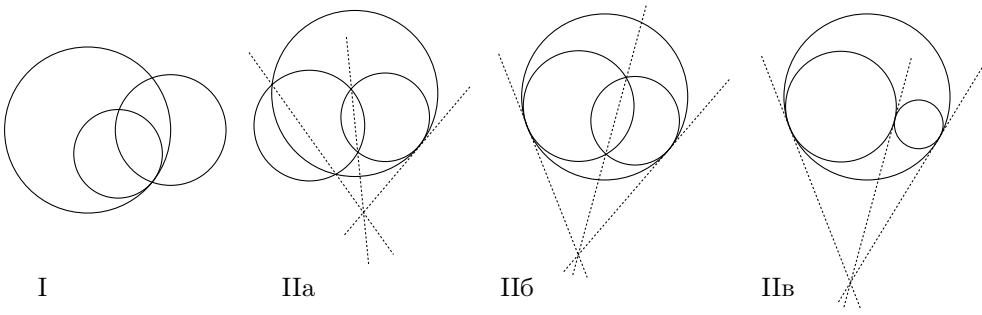


Рис. 17. Случаи касания двух окружностей

окружности в прямые линии; поэтому две стороны треугольника становятся параллельными прямыми. В случае II все варианты — гиперболические, одна или больше вершин треугольника лежат на бесконечно удалённой окружности (или прямой). (Вопрос: *чему равен угол, вершина которого лежит на бесконечно удалённой окружности*?) Отметим, что среди предельных случаев нет сферических.

Изменение точки зрения: подход с использованием преобразований

Мы признательны Ки́ту Бёрнсу, указавшему на следующую весьма изящную формулировку и доказательство основной теоремы. Его доказательство использует группу преобразований Мёбиуса пополненной комплексной плоскости (т. е. плоскости, рассматриваемой как поле комплексных чисел с добавлением бесконечно удалённой точки). Преобразования Мёбиуса обратимы, непрерывны в обе стороны, конформны и переводят окружности в широком смысле (включая прямые) в окружности. Далее, если даны две упорядоченные тройки точек, то существует (единственное) преобразование Мёбиуса, переводящее каждую точку первой тройки в соответствующую точку второй. (В прекрасной, очень понятно написанной книге о геометрии с клейновской точки зрения [4] подробно изучается группа Мёбиуса и её геометрические свойства, см. гл. 5.)

В частности, и это всё, что нам потребуется, найдётся преобразование Мёбиуса, переводящее любую заданную точку в бесконечно удалённую. Как мы уже видели, преобразование с нужными свойствами можно построить как композицию двух стереографических проекций.

Рассмотрим криволинейный треугольник ABC, образованный тремя окружностями c_1, c_2 и c_3. Без потери общности можно считать, что точка A лежит на c_1 и c_2, и пусть A' — вторая точка пересечения этих окружностей. Тогда возможны три расположения точек A и A' относительно третьей окружности c_3, которые соответствуют трём случаям в лемме и теореме:

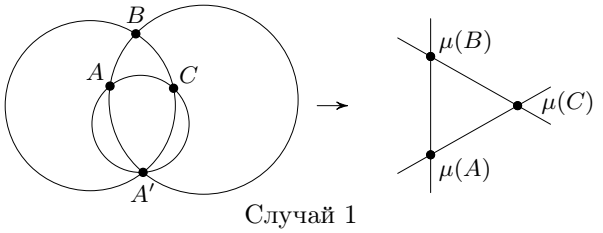
A' лежит на c_3;
A' и A лежат по разные стороны от c_3;
A' и A лежат по одну сторону от c_3.

Простым изменением точки зрения, а именно, поместив A' в бесконечность, мы получаем возможность наглядно сопоставить сумму углов треугольника ABC с 180°. Хотя такой подход не выявляет соответствующую глобальную геометрию, в которой окружности являются геодезическими, он обладает преимуществом замечательной простоты и непосредственности.

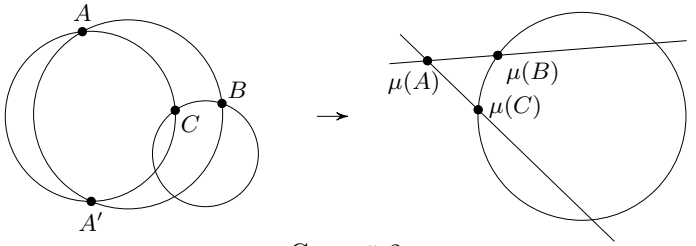
Теорема (альтернативная формулировка). *Пусть c_1, c_2 и c_3 — три окружности на плоскости, причём каждые две пересекаются в двух различных точках. Тогда выполнено ровно одно из следующих трёх условий, которые определяют сумму углов любого треугольника, образованного тремя окружностями*:

1) *пересечение каждой пары окружностей содержит точку третьей окружности, и в этом случае сумма углов равна* $180°$;
2) *пересечение каждой пары окружностей лежит целиком внутри или вне третьей окружности, и в этом случае сумма углов меньше* $180°$;
3) *в пересечении каждой пары окружностей одна точка лежит внутри, а другая вне третьей окружности, и в этом случае сумма углов больше* $180°$.

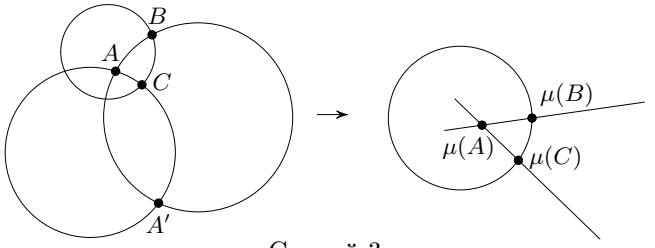
Доказательство. Пусть A, B и C — вершины треугольника, образованного окружностями c_1, c_2 и c_3, причём $\{A, A'\} = c_1 \cap c_2$. Сумма углов треугольника не зависит от того, какую пару окружностей мы здесь



Случай 1



Случай 2



Случай 3

Рис. 18. Действие преобразования Мёбиуса μ, переводящего A' в ∞

рассматриваем, поэтому достаточно показать, что эта сумма определяется лишь расположением точек A и A' относительно окружности c_3.

Применим преобразование Мёбиуса μ, переводящее A' в ∞. При этом преобразовании окружности c_1 и c_2, проходящие через A', переходят в прямые.

Если A' лежит на c_3, то эта окружность тоже переходит в прямую, поэтому сумма углов треугольника ABC равна $180°$.

Если A и A' лежат по одну сторону от c_3, то образ окружности c_3 — окружность, причём образ точки A лежит *вне* её, поскольку ∞ лежит вне её; следовательно, сумма углов образа треугольника ABC меньше $180°$.

Наконец, если A и A' лежат по разные стороны от c_3, то образ окружности c_3 — снова окружность, причём образ точки A лежит *внутри* неё; следовательно, сумма углов образа треугольника ABC больше $180°$ (рис. 18). □

Заключение

Все классические геометрии — евклидова, сферическая и гиперболическая — имеют множество конформных реализаций на плоскости, получаемых посредством стереографической проекции. Ознакомившись с этими реализациями, мы можем полностью классифицировать треугольники, стороны которых — отрезки или дуги окружностей, и в соответствии с этой классификацией оценить сумму их углов. Наконец, используя группу конформных преобразований пополненной плоскости, мы смогли усовершенствовать наше решение, сделав его очень простым, хотя и менее информативным. Наш успех указывает на важность изучения неевклидовых геометрий, а также групп преобразований.

Благодарности

Кроме Ки́та Бёрнса, мы благодарим Дуэйна Бролайна, Лео Камерфорда, Боба Фута, Юрия Ионина и Розмари Шмальц за полезные комментарии. Мы хотели бы также поблагодарить нашего первого рецензента, сообщившего нам рассуждение, приведённое в замечании 3.

Список литературы

[1] *Гильберт Д., Кон-Фоссен С.* Наглядная геометрия. М.: Наука, 1981.

[2] *Дубровин Б. А., Новиков С. П., Фоменко А. Т.* Современная геометрия: Методы и приложения. Том I. Геометрия поверхностей, групп преобразований и полей. М.: Эдиториал УРСС, 1998.

[3] *Коксетер Г. С. М., Грейтцер С. Л.* Новые встречи с геометрией. Ижевск: РХД, 2002.

[4] *Brannan D., Esplen M., Gray J.* Geometry. Cambridge: Cambridge University Press, 1999.

[5] *Greenberg M. J.* Euclidean and non-Euclidean geometries. New York: Freeman, 1993.

[6] *McCleary J.* Geometry from a differentiable viewpoint. Cambridge: Cambridge University Press, 1994.

[7] *Moise E.* Elementary geometry from an advanced standpoint. MA: Addison-Wesley, Reading, 1990.

[8] *O'Neill B.* Semi-Riemannian geometry. With applications to relativity. New York: Academic Press, 1983. (Pure and Applied Math.; V. 103).

Г. Гальперин, Восточно-Иллинойсский университет, Чарлстон, США
ggalperin@eiu.edu

Ч. Делман, Восточно-Иллинойсский университет, Чарлстон, США
cidelman@eiu.edu

Наш семинар: математические сюжеты

Средние Джини

А. Б. Певный, С. М. Ситник

1. Широко известны три вида средних значений для пары положительных чисел a, b. Это *среднее арифметическое* $A(a,b)$, *среднее геометрическое* $G(a,b)$ и *среднее гармоническое* $H(a,b)$, которые определяются по формулам

$$A(a,b) = \frac{a+b}{2}, \quad G(a,b) = \sqrt{ab}, \quad H(a,b) = \frac{2ab}{a+b}.$$

Эти средние были введены ещё в Древней Греции в работах Пифагора и его школы, а также Никомаха [4, 15]. Источником введения средних для древнегреческих учёных стали пропорции, учение о которых активно развивалось, так как понятие пропорции использовалось тогда не только в математике, но также и в философии, скульптуре, архитектуре и живописи [4, 15]. Например, если найти величину x из нижеследующих пропорций, то как раз получим последовательно основные средние:

$$\begin{aligned} \frac{a-x}{x-b} = \frac{a}{a} = 1 \quad &\Rightarrow \quad x = A(a,b), \\ \frac{a-x}{x-b} = \frac{a}{x} \quad &\Rightarrow \quad x = G(a,b), \\ \frac{a-x}{x-b} = \frac{a}{b} \quad &\Rightarrow \quad x = H(a,b). \end{aligned}$$

Позднее были введены *среднее квадратичное*

$$Q(a,b) = \sqrt{\frac{a^2+b^2}{2}}$$

Математическое просвещение, сер. 3, вып. 20, 2016(135–142)

и общие *степенные средние порядка* r, которые для набора неотрицательных (при $r \geqslant 0$) или положительных (при $r < 0$) чисел определяются по формулам

$$\begin{gathered} M_r(a_1, a_2, \ldots, a_n) = \left(\frac{a_1^r + a_2^r + \ldots + a_n^r}{n}\right)^{1/r}, \\ M_0(a_1, a_2, \ldots, a_n) = \lim_{r \to 0} M_r(a_1, a_2, \ldots, a_n) = \sqrt[n]{a_1 \cdot a_2 \cdot \ldots \cdot a_n}. \end{gathered} \tag{1}$$

При этом справедливы формулы, выражающие основные средние через степенные. Например, для двух чисел

$$\begin{aligned} M_{-1}(a, b) &= H(a, b), \quad M_0(a, b) = G(a, b), \\ M_1(a, b) &= A(a, b), \quad M_2(a, b) = Q(a, b). \end{aligned}$$

Важнейшим свойством степенных средних является тот факт, что семейство этих величин монотонно возрастает по параметру r, или, как говорят, образует шкалу [1, 7, 10–12]. В частности,

$$H(a, b) \leqslant G(a, b) \leqslant A(a, b).$$

Эта самая известная из целого ряда теорем, в которой сравниваются между собой различные средние.

Наряду со степенными в различных задачах потребовалось использовать и другие типы средних [7, 11, 14]. В 1938 году итальянский математик и экономист Коррадо Джини (Corrado Gini, 1884–1965) ввёл новое семейство средних, которые впоследствии были названы его именем. Эти средние величины определяются так. Для любого набора положительных чисел $a = (a_1, a_2, \ldots, a_n)$ введём степенные суммы

$$S_p(a) = \sum_{i=1}^{n} a_i^p. \tag{2}$$

Тогда *средние Джини* $Gi_{p,q}(a)$, зависящие от двух вещественных параметров p и q, определяются по формулам

$$Gi_{p,q}(a) = Gi_{q,p}(a) = \begin{cases} \left(\dfrac{S_p(a)}{S_q(a)}\right)^{\frac{1}{p-q}}, & \text{если } p \neq q; \\ \exp\left(\dfrac{\sum_{i=1}^n a_i^p \ln a_i}{\sum_{i=1}^n a_i^p}\right), & \text{если } p = q \neq 0; \\ M_0(a), & \text{если } p = q = 0. \end{cases} \tag{3}$$

При этом нетрудно показать, что в соотношениях (3) второй случай получается из общего первого, а третий из второго при помощи предельных переходов. В частном случае, когда один из параметров равен нулю,

среднее Джини сводится к степенному среднему (1), а именно, справедливо соотношение

$$Gi_{p,0}(a_1, a_2, \ldots, a_n) = M_p(a_1, a_2, \ldots, a_n). \tag{4}$$

Мы далее рассматриваем общий случай, положив для определённости $p > q$.

Основным предметом изучения Коррадо Джини в экономической теории было имущественное неравенство, для его описания он ввёл знаменитый «*индекс Джини*», для вычисления и анализа которого использовались в том числе и различные классы средних. Это одно из основных понятий современной социальной статистики [8], а сам К. Джини — один из общепризнанных её создателей. Его результаты также известны в области статистического анализа проблем демографии. Джини использовал и пропагандировал результаты своего соотечественника Вильфредо Парето. Парето первым провёл конкретные расчёты, из которых следовало, что в его время всего 20 % населения владело 80 % национального богатства, и сделал вывод, что организованное таким образом государство ждёт быстрая и неизбежная гибель. Отметим, что сейчас в России реальные показатели индексов Джини ещё хуже (для неимущих). Развивая работы Парето, американский экономист М. О. Лоренц разработал теорию оценки разницы доходов населения по группам (кривая Лоренца), которая сейчас называется методологией Парето — Лоренца — Джини [8]. Отметим и третьего знаменитого итальянского математика и статистика XX века — Бруно де Финетти, разработавшего собственную оригинальную систему взглядов на основания теории вероятностей, автора известного в финансовой математике *парадокса де Финетти* (если капитал страховой компании конечен, то за бесконечное время она наверняка разоряется).

Несмотря на то, что по запросу в интернете на фамилии Парето, Джини или де Финетти будет выдано несколько тысяч ссылок, найти информацию о них самих практически невозможно. Это неумное замалчивание есть проявление так называемой политкорректности и связано с тем, что все они были сторонниками итальянского фашизма, в котором, вероятно, видели способ возрождения Италии. На идеях Парето воспитывался Муссолини, а Джини при нём занимал высокие посты и получал полную поддержку. Де Финетти прошёл путь от идейного сторонника фашизма до убеждённого коммуниста, читавшего лекции вьетнамским бойцам в джунглях под американскими бомбёжками. Разумеется, идеология фашизма во всех его проявлениях достойна однозначного осуждения, как и все её носители, кем бы сами они себя ни считали. Однако это не повод замалчивать значительный вклад этих учёных в общечеловеческую культуру и науку, тем более что они никак не запятнали себя непосредственным участием в преступлениях. Понятие оптимума по Парето, средних Джини и индекса

Джини, теория Колмогорова — Нагумо — де Финетти квазиарифметических средних, названная по именам её создателей, — всё это важнейшие и необходимые составляющие современной математики и статистики.

Важный частный случай средних Джини получается из (3) при $q = p - 1$. Эти средние были переоткрыты в 1971 г. Д. Лемером и имеют вид

$$Le_p(a) = \frac{\sum_{i=1}^{n} a_i^p}{\sum_{i=1}^{n} a_i^{p-1}}.$$

Их в советское время чуть позже Лемера также переоткрыл школьный учитель Ю. М. Фирсов [5].

2. Покажем, что средние Джини, так же как и степенные средние, образуют шкалу по своим параметрам, для них справедлива теорема сравнения. А именно, средние Джини увеличиваются, когда или параметр p, или параметр q, или оба эти параметра увеличиваются.

Теорема. *Пусть среди положительных чисел $(a_1, a_2, \ldots, a_n)$ есть хотя бы два неравных числа. Если выполнены условия $p_1 > q_1$, $p_2 > q_2$, $p_2 \geqslant p_1$, $q_2 \geqslant q_1$ и хотя бы одно из двух последних неравенств строгое, то справедливо неравенство*

$$Gi_{p_1,q_1}(a) < Gi_{p_2,q_2}(a). \tag{5}$$

Доказательство. Неравенство (5) в развёрнутом виде, записанное через степенные суммы (2), выглядит так:

$$\left(\frac{S_{p_1}(a)}{S_{q_1}(a)}\right)^{\frac{1}{p_1-q_1}} < \left(\frac{S_{p_2}(a)}{S_{q_2}(a)}\right)^{\frac{1}{p_2-q_2}}.$$

Оно упрощается, если от обеих частей взять логарифм:

$$\frac{1}{p_1 - q_1}(\ln S_{p_1}(a) - \ln S_{q_1}(a)) < \frac{1}{p_2 - q_2}(\ln S_{p_2}(a) - \ln S_{q_2}(a)). \tag{6}$$

Введём функцию $f(p) = \ln S_p(a)$. Тогда неравенство (6) переписывается так:

$$\frac{f(p_1) - f(q_1)}{p_1 - q_1} < \frac{f(p_2) - f(q_2)}{p_2 - q_2}. \tag{7}$$

Но (7) заведомо выполняется, если функция $f(p)$ строго выпукла. Действительно (см. рис. 1), левая часть (7) — это тангенс наклона секущей, проходящей через точки $(p_1, f(p_1))$, $(q_1, f(q_1))$, а правая часть — это тангенс наклона второй секущей. В условиях теоремы первый тангенс меньше второго тангенса (см. рис. 1).

Осталось установить строгую выпуклость функции $f(p)$, для чего достаточно установить неравенство $f''(p) > 0$, $-\infty < p < \infty$. Вычисление даёт

$$f''(p) = \frac{1}{S^2(p)}(S''(p)S(p) - (S'(p))^2).$$

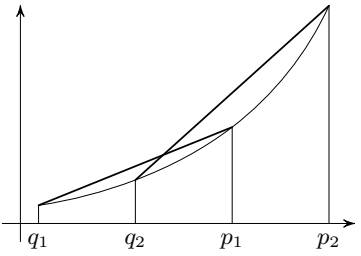


Рис. 1

Это выражение будет положительным, если выполнено условие

$$(S'(p))^2 < S''(p)S(p). \tag{8}$$

В развёрнутом виде неравенство (8) эквивалентно следующему:

$$\left(\sum_{i=1}^{n} a_i^p \ln a_i\right)^2 < \sum_{i=1}^{n} a_i^p \sum_{i=1}^{n} a_i^p (\ln a_i)^2.$$

Но последнее соотношение следует из неравенства Коши — Буняковского для векторов с координатами $\sqrt{a_i^p}$ и $\sqrt{a_i^p}\ln a_i$.

Покажем, что действительно неравенство (8) всегда является строгим, хотя в использованном нами неравенстве Коши — Буняковского возможен и случай равенства. В этом случае должно существовать такое число λ, что

$$\sqrt{a_i^p}\ln a_i = \lambda\sqrt{a_i^p}, \quad i = 1, 2, \ldots, n,$$

а это возможно, только если все числа a_i равны, что противоречит условию теоремы. Следовательно, в (8) всегда будет строгое неравенство. Поэтому при сформулированных в условиях теоремы предположениях также строгими будут неравенства (7) и (5).

Это завершает доказательство теоремы. □

По мнению авторов, приведённое доказательство проще, чем в монографии Буллена [11, с. 249], в том числе и потому, что не содержит ссылок на результаты других работ.

В качестве следствия рассмотрим неравенства между средними Джини и более известными степенными средними, которые получаются с учётом соотношения (4).

Следствие. *Пусть среди положительных чисел $a_1, a_2, \ldots, a_n$ есть хотя бы два неравных. Тогда:*

1) *если $r \geqslant p > q$, $r > 0 > q$, то справедливо неравенство*

$$Gi_{p,q}(a_1, a_2, \ldots, a_n) < M_r(a_1, a_2, \ldots, a_n);$$

2) *если $p \geqslant r > 0$, $p > q > 0$, то справедливо неравенство*

$$M_r(a_1, a_2, \ldots, a_n) < Gi_{p,q}(a_1, a_2, \ldots, a_n).$$

В качестве упражнения предлагаем читателю самостоятельно доказать, используя определение (3), что выполнены соотношения

$$\lim_{p\to\infty} Gi_{p,q}(a) = \max_{1\leqslant i\leqslant n} a_i, \quad \lim_{q\to-\infty} Gi_{p,q}(a) = \min_{1\leqslant i\leqslant n} a_i,$$

из которых следует, что среднее Джини (с учётом доказанной теоремы о монотонности по параметрам!) находится между минимальным и максимальным из чисел a_i:

$$\min_{1\leqslant i\leqslant n} a_i \leqslant Gi_{p,q}(a) \leqslant \max_{1\leqslant i\leqslant n} a_i,$$

и, таким образом, является настоящим средним.

3. Средние значения достаточно широко используются в химии, в частности в теории полимеризации. В качестве примеров укажем известные монографии [2, 3, 9]. Так, в монографии [2] в главе о молекулярно-массовом распределении изучаются различные способы усреднения по массе молекул и по числу частиц. Вводится и изучается весовое среднее арифметическое [2, с. 127], для средневязкостной молекулярной массы вводится весовое степенное среднее вида (1), см. [2, с. 130]. С математической точки зрения, при этом без доказательства используется монотонность весовых степенных средних по параметру (свойство шкалы средних, см. [1, 7, 10–12]), а также неравенство между весовыми средними и невесовым средним арифметическим. При изучении связей механизмов полимеризации с молекулярно-массовым распределением вводятся и изучаются свойства интегральных аналогов средних Джини $Gi_{3,2}, Gi_{2,1}$ [2, с. 184–185], с их использованием определяется так называемый коэффициент полидисперсности Шульца.

В монографии [3] при изучении физических методов исследования макромолекул в растворах вводится так называемое вискозиметрическое среднее, которое выражается через молекулярные веса макромолекул с помощью среднего Джини $Gi_{s+1,1}$ [3, с. 36].

Различные средние величины также широко используются в известной монографии [9]. Так, на с. 23–24 для средних значений молекулярных весов вводятся последовательно весовое интегральное среднее произвольного порядка, вес которого равен численной функции распределения; среднее Джини, совпадающее со средним Лемера $Gi_{q,q-1} = Le_q$, которое называется q-средним молекулярным весом, а также степенное среднее. На с. 67 при статистическом анализе механизмов фракционирования вводится среднее Джини — Лемера $Gi_{2,1} = Le_2$. На с. 83 при изучении гидродинамических средних весов и связанных с ними критериев полидисперсности вводятся понятия среднедиффузионного и среднеседиментационного весов, которые выражаются через весовые степенные интегральные средние некоторых

отрицательных порядков, а также весовое интегральное среднее Джини $Gi_{1,1-b}$, где b — это некоторая характеристика, изменяющаяся в пределах $0 < b < 1$. При этом без строгого доказательства, с необычной для математика формулировкой «обычно это верно» утверждается, что указанное среднее Джини всегда больше среднего арифметического. Отметим, что строгое доказательство этого утверждения вытекает из приведённого выше следствия при выборе параметров $p = 1$, $q = 1 - b$, $0 < q < 1$, $r = 1$. На с. 83–84 вводится некоторая величина, связанная с молекулярными весами, которая выражается через среднее Джини — Лемера $Gi_{2-b,1-b} = Le_{2-b}$, а на с. 229 для минимальной величины эффективного параметра получено выражение через среднее Джини $Gi_{3/2,-3/2}$. Вопросы предельного поведения средних Джини в терминах отношения степенных средних рассматриваются также на с. 252 при изучении возможностей сведéния процессов полимеризации к основным статистическим классам распределений.

Таким образом, различные классы средних значений, включая средние Джини, находят важные применения в химии при рассмотрении многочисленных вопросов теории полимеризации молекул. При этом ряд используемых в химической литературе нестрогих соотношений между средними может быть строго обоснован с использованием доказанных в настоящей статье неравенств между средними Джини и степенными средними.

4. В заключение отметим, что так как в нашем доказательстве используется неравенство Коши — Буняковского, то можно получить некоторые усиления доказанных неравенств с использованием уточнений неравенства Коши — Буняковского, полученных в [7, 14]. Авторы также использовали неравенства для средних Джини при доказательстве интересного неравенства А. С. Гаспаряна между однородными степенными формами [6, 13].

Список литературы

[1] *Беккенбах Э., Беллман Р.* Неравенства. М.: Мир, 1965.

[2] *Берлин Ал. Ал., Вольфсон С. А., Ениколопян Н. С.* Кинетика полимеризационных процессов. М.: Химия, 1978.

[3] *Волькенштейн М. В.* Конфигурационная статистика полимерных цепей. М.-Л.: АН СССР, 1959.

[4] *Джини К.* Средние величины. М.: Статистика. 1970.

[5] *Калинин С. И.* Средние величины степенного типа. Неравенства Коши и Ки Фана. Киров: Изд. ВятГГУ, 2002.

[6] *Певный А. Б., Ситник С. М.* Об одном неравенстве А. С. Гаспаряна // Проблемы матем. анализа. 2014. Вып. 77. С. 159–162.

[7] *Ситник С. М.* Уточнения и обобщения классических неравенств // Итоги науки. ЮФО Серия «Математический форум: Исследования по математическому анализу», ред. Ю. Ф. Коробейник, А. Г. Кусраев. Владикавказ, 2009. Т. 3. С. 221–266.

[8] Социальная статистика / Ред. И. И. Елисеева. М.: Финансы и статистика, 2003.

[9] *Френкель С. Я.* Введение в статистическую теорию полимеризации. М.-Л.: Наука, 1965.

[10] *Харди Г. Г., Литтльвуд Дж. Е., Полиа Г.* Неравенства. М.: ИЛ, 1948.

[11] *Bullen P. S.* Handbook of means and their inequalities. Dordrecht: Kluwer Academic Publishers, 2003. (Mathematics and its Applications (East European Series); V. 560).

[12] *Mitrinović D. S., Pečarić J. E., Fink A. M.* Classical and new inequalities in analysis. Dordrecht: Kluwer Academic Publishers, 1993. (Mathematics and its Applications (East European Series); V. 61).

[13] *Pevnyi A. B., Sitnik S. M.* On Gasparyan's inequality // J. Math. Sci. 2015. V. 205, № 2. P. 304–307.

[14] *Sitnik S. M.* Generalized Young and Cauchy — Bunyakowsky inequalities with applications: a survey. http://arxiv.org/abs/1012.3864.

[15] *Toader Gh., Toader S.* Greek means and the arithmetic–geometric mean. RGMIA Monographs, Victoria University, 2005. http://rgmia.org/monographs/toader.htm

А. Б. Певный, Сыктывкарский государственный университет
pevnyi@syktsu.ru

С. М. Ситник, Воронежский институт МВД России
pochtasms@gmail.com

Гиперболические многочлены и конусы Гординга

Н. В. Филимоненкова, П. А. Бакусов

В статье представлен краткий обзор теории гиперболических многочленов, созданной шведским математиком Ларсом Гординг-ом в середине XX века. Кратко описано происхождение этой теории. Приведены основные факты о гиперболических многочленах и связанных с ними конусах Гординга, указаны типичные примеры и области приложения, в особенности та роль, которую они сыграли в современной теории полностью нелинейных дифференциальных уравнений в частных производных.

§ 1. Введение

Изучение алгебраических многочленов имеет древнюю историю, в частности изучение распределения корней многочлена, количества вещественных корней. Гиперболический многочлен — это современное наименование многочленов, у которых все корни вещественные.

Определение 1.1. Многочлен одной переменной $p(t)$ называется гиперболическим, если он имеет только вещественные корни.

Для многочлена нескольких переменных вещественность его корней можно описать путём сведения к многочлену одной переменной. Например, для однородных многочленов это делается следующим образом.

Определение 1.2. Однородный многочлен нескольких переменных $P(x)$, $x = (x_1, x_2, \dots, x_n)$, называется a-гиперболическим или гиперболическим в направлении вектора $a \in \mathbb{R}^n$, если $P(a) \neq 0$ и многочлен одной переменной $p(t) = P(at + x)$ имеет только вещественные корни для любого вектора $x \in \mathbb{R}^n$.

Работа выполнена при поддержке РФФИ, проекты № 15–01–07650 а, № 15–31–20600 мол-а-вед.

Употребление эпитета «гиперболический» в приложении к многочленам становится понятно после выяснения первоисточника определения 1.2. По всей видимости, уже после того как было создано понятие a-гиперболического многочлена нескольких переменных, этот эпитет распространился и на многочлены одной переменной, у которых все корни вещественные.

Понятие a-гиперболического многочлена нескольких переменных восходит к шведскому математику Ларсу Гордингу (Lars Gårding, 1919–2014), который заимствовал некоторые идеи из работ И. Г. Петровского (к примеру, Гординг нередко ссылается на статью Петровского [8]). Корни этого понятия принадлежат теории линейных дифференциальных уравнений в частных производных (LPDE — linear partial differential equations), а точнее теории гиперболических уравнений порядка m. Гиперболические уравнения высшего порядка определяются путём выделения одного из ключевых свойств гиперболических уравнений второго порядка и обобщения его на случай $m > 2$. Одним из таких ключевых свойств является устойчивость решения задачи Коши (непрерывная зависимость решения от начальных данных). В работе [19] Гординг показал, как это свойство можно выразить в терминах характеристического многочлена (символа) дифференциального оператора. Для этого он ввёл понятие a-гиперболического многочлена и доказал, что решение задачи Коши в n-мерном полупространстве, ограниченном гиперплоскостью $\{x \in \mathbb{R}^n \colon (x, a) = 0\}$, устойчиво тогда и только тогда, когда характеристический многочлен оператора является a-гиперболическим. Далее этот критерий удалось свести к исследованию главной части оператора, т. е. к однородному многочлену. Генезис понятия гиперболического многочлена и его связь с понятием устойчивости подробно описаны в статье [35].

Поскольку самым знаменитым представителем гиперболических операторов является волновой оператор

$$F[u] = \frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2} - \frac{\partial^2 u}{\partial x_3^2} - \ldots - \frac{\partial^2 u}{\partial x_n^2},$$

то прародителем всех a-гиперболических многочленов можно считать его символ:

$$P(x) = x_1^2 - x_2^2 - x_3^2 - \ldots - x_n^2.$$

В 1959 году Гординг написал небольшую статью [18], в которой исследовал a-гиперболические многочлены с алгебраической точки зрения, независимо от теории гиперболических LPDE, и вывел ряд фундаментальных фактов. Статья [18] не вызвала немедленного резонанса: она казалась скромной как по объёму (всего десять страниц), так и по значению

описанных в ней результатов. Долгое время теория a-гиперболических многочленов распространялась в сравнительно небольшом кругу математиков (см. [11, 15, 23, 31, 34, 38]) и рассматривалась в контексте гиперболических LPDE.

Так продолжалось до тех пор, пока теория Гординга не обрела неожиданные приложения в теории полностью нелинейных дифференциальных уравнений в частных производных (FNPDE — fully nonlinear partial differential equations). В первую очередь следует назвать фундаментальную работу [14] 1985 года, авторы которой L. Caffarelly, L. Nirenberg, J. Spruck впервые в математической литературе упоминают статью Гординга [18] в связи с описанием корректного множества разрешимости уравнений типа Монжа — Ампера.

В начале XXI века к работе Гординга [18] пришло наконец заслуженное признание, и в настоящее время теория a-гиперболических многочленов и связанных с ними конусов Гординга востребована сразу в нескольких областях математики.

Во-первых, это уже названная современная теория FNPDE, о которой будет подробнее рассказано в § 5. После публикации [14] гиперболические многочлены и конусы Гординга стали генератором множества алгебро-геометрических понятий в нелинейной теории, которые нуждаются в постоянном пересмотре, уточнении, систематизации. Исследованием этого фундамента занимаются F. R. Harvey, H. B. Lawson [22] и в большей степени Н. М. Ивочкина, Н. В. Филимоненкова [5, 26, 27].

Во-вторых, в конце 1990-х годов сформировалась новая ветвь линейного программирования, названная гиперболическим программированием (hyperbolic programming), основоположником которого считается O. Guler [12, 20, 37]. Задача оптимизации линейного функционала ставится для конуса Гординга.

В-третьих, есть работы, посвящённые алгебраическим свойствам a-гиперболических многочленов. В частности, исследуется возможность представления гиперболических многочленов разной степени в виде определителя (determinantal representability) [13, 30]. Эта ветка исследований смыкается с приложениями гиперболических многочленов в комбинаторике и дискретной теории вероятности. В этой области работают, например, P. Branden, R. Pemantle [35], L. Gurvits [21].

В-четвёртых, отдельное направление составляют работы, посвящённые гиперболическим многочленам одной переменной, хотя, разумеется, такие многочлены изучались задолго до того, как появилась теория Гординга. Этому направлению принадлежит работы таких математиков, как В. И. Арнольд [1], В. Костов [29], F. Rellich [36], A. Rainer и др. Исследуются

распределение корней гиперболического многочлена, геометрия области гиперболических многочленов фиксированной степени, порядок гладкости корней многочлена в зависимости от возмущения его коэффициентов, критерии гиперболичности.

Надо заметить, что большинство работ, так или иначе связанных с a-гиперболическими многочленами, принадлежат зарубежным авторам. Насколько нам известно, в отечественной математике эту теорию активно использует только Н. М. Ивочкина, специалист по FNPDE, а также её ученики, к которым относятся и авторы данной статьи.

Таким образом, первой целью предлагаемой работы является популяризация теории a-гиперболических многочленов Гординга, знакомство с ней российских читателей. Второй целью работы является демонстрация «простых» задач этой теории, доступных даже студентам математических и технических специальностей. В конце концов, если не вдаваться в глубину приложений, то понятие a-гиперболического многочлена — это просто удобный способ исследования вещественности корней для многочлена нескольких переменных. Поэтому анализ его алгебраических свойств и построение примеров относятся к области элементарной математики, доступной широкой аудитории.

§ 2. Гиперболические многочлены одной переменной

В этом параграфе представляем читателю фрагмент классической истории изучения многочленов одной переменной, у которых все корни вещественные и которые позднее стали называться гиперболическими. Приводим обзор известных неравенств Ньютона и Маклорена, указываем историю их появления и возможность использования в разных формах.

Рассмотрим комплексный многочлен одной переменной

$$p(t) = \sum_{k=0}^{n} a_k t^{n-k}$$

степени n. Согласно определению 1.1 он называется гиперболическим, если имеет n вещественных корней. В этом случае

$$p(t) = a_0 \prod_{k=1}^{n} (t - t_k), \quad t_k \in \mathbb{R},$$

а значит, $p(t)/a_0$ — многочлен с вещественными коэффициентами. Поэтому, не умаляя общности, далее рассматриваем вещественные гиперболические многочлены.

Теорема 2.1. *Если* $p(t) = \sum_{k=0}^{n} a_k t^{n-k}$ *— гиперболический многочлен, то все его коэффициенты удовлетворяют неравенствам Ньютона*:

$$\frac{a_{k-1}}{C_n^{k-1}} \cdot \frac{a_{k+1}}{C_n^{k+1}} \leqslant \left(\frac{a_k}{C_n^k}\right)^2, \quad 1 \leqslant k \leqslant n-1. \tag{2.1}$$

Если первые m *коэффициентов многочлена положительны, то эти коэффициенты удовлетворяют неравенствам Маклорена*:

$$\frac{a_1}{C_n^1} \geqslant \left(\frac{a_2}{C_n^2}\right)^{1/2} \geqslant \left(\frac{a_3}{C_n^3}\right)^{1/3} \geqslant \ldots \geqslant \left(\frac{a_{m-1}}{C_n^{m-1}}\right)^{1/(m-1)} \geqslant \left(\frac{a_m}{C_n^m}\right)^{1/m}. \tag{2.2}$$

Неравенства (2.1) получили своё название в честь Исаака Ньютона, поскольку они восходят к правилу определения числа «возможных» (вещественных) и «невозможных» (комплексных) корней многочлена, впервые изложенному в книге Ньютона [33]. После переложения на современный математический язык идей Ньютона получается, что по существу он утверждал следующее. Количество комплексных корней не меньше, чем число перемен знака в цепочке разностей

$$\left(\frac{a_k}{C_n^k}\right)^2 - \frac{a_{k-1}}{C_n^{k-1}} \cdot \frac{a_{k+1}}{C_n^{k+1}}, \quad 1 \leqslant k \leqslant n-1.$$

Если верить примечаниям к русскому переводу [7] книги [33], то остаётся неизвестным, как Ньютон открыл это правило, доказал же его много позже J. J. Sylvester. Неравенства Ньютона, которые являются частным случаем этого правила, имеют несколько методов доказательства. Наиболее простой метод содержится в четвёртой главе монографии G. H. Hardy, J. E. Littlewood, G. Polya [10]: после «расщепления» переменной t на две переменные и дифференцирования многочлен степени n приводится к квадратному многочлену, вещественность корней которого обусловлена теоремой Ролля и равносильна неотрицательности дискриминанта. Воспроизведение этого доказательства представляется несложной учебной задачей. Ещё одной простой задачей является диагностика двух случаев, когда в цепочке (2.1) часть или все неравенства вырождаются в равенства. В математической литературе один из двух случаев часто бывает упущен.

Замечание 1. Неравенства (2.1) являются необходимым, но не достаточным условием гиперболичности многочлена. Действительно, можно построить многочлен любой степени, для коэффициентов которого выполнены неравенства Ньютона, но не все его корни вещественны. Например, $p(t) = 2t^3 + 6t^2 + 6t + 1$.

Неравенства (2.2) впервые получил C. MacLaurin в работе [32], где он использовал геометрический метод, основанный на делении отрезка и мак-

симизации суммы произведений получившихся частей. При этом длины частей отрезка играли роли корней многочлена, а длина целого отрезка — их суммы, следовательно, метод Маклорена применим к многочленам с корнями одинакового знака. Возможно, именно поэтому до сей поры неравенства (2.2) принято связывать только с такими гиперболическими многочленами, у которых все корни отрицательны или, что то же самое, все коэффициенты положительны (в этом случае цепочка (2.2) содержит максимальное число неравенств).

Однако во второй главе книги [10] приведено чисто арифметическое доказательство неравенств Маклорена, основанное на неравенствах Ньютона: нужно рассмотреть первые несколько неравенств (2.1), возвести k-е неравенство в степень k, перемножить и сократить повторяющиеся множители. Для успешной реализации этого метода достаточно, чтобы первые несколько коэффициентов многочлена были положительны. Причём количество неравенств в цепочке (2.2) соответствует количеству положительных коэффициентов.

Неравенства Ньютона и Маклорена могут быть записаны в трёх разных формах за счёт следующей подстановки (см. ниже определение 2.2):

$$a_k = \sigma_k(\lambda) = \operatorname{tr}_k S. \tag{2.3}$$

Здесь $\sigma_k(\lambda)$ — элементарная симметрическая функция от вещественных переменных $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$:

$$\sigma_k(\lambda) = \sum_{i_1, i_2, \ldots, i_k} \lambda_{i_1}\lambda_{i_2}\ldots\lambda_{i_k}, \quad \sigma_0(\lambda) = 1.$$

Первое равенство в формуле (2.3) объясняется тем, что по любому вектору λ можно построить гиперболический многочлен $p(t) = \prod_{k=1}^{n}(t + \lambda_k)$ с корнями $-\lambda_k$ и коэффициентами $a_k = \sigma_k(\lambda)$ (теорема Виета). В частности, $a_1 = \lambda_1 + \lambda_2 + \ldots + \lambda_n$, $a_n = \lambda_1\lambda_2\ldots\lambda_n$, и если все λ_k положительны, то крайние члены цепочки (2.2) дают известное неравенство между средним арифметическим и средним геометрическим:

$$\frac{\lambda_1 + \lambda_2 + \ldots + \lambda_n}{n} \geqslant \sqrt[n]{\lambda_1\lambda_2\ldots\lambda_n}.$$

Поясним второе равенство в формуле (2.3). Обозначим символом $\operatorname{Sym}(n)$ пространство всех симметричных матриц размера $n \times n$.

Определение 2.2. Главный минор порядка m матрицы $S \in \operatorname{Sym}(n)$ — это определитель матрицы, которая составлена из элементов матрицы S, стоящих на пересечении каких-либо m строк и m столбцов (с одинаковыми

номерами). Следом порядка k или просто k-следом матрицы S называется сумма всех главных миноров порядка k матрицы S. Обозначение: $\operatorname{tr}_k S$. По определению $\operatorname{tr}_0 S = 1$.

Понятие k-следа включает два крайних случая: 1-след — это просто след матрицы, сумма элементов на главной диагонали, $\operatorname{tr}_1 S = \operatorname{tr} S$, n-след — это определитель матрицы, $\operatorname{tr}_n S = \det S$. К примеру, 2-след матрицы размера 3×3 вычисляется следующим образом:

$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix},$$

$$\operatorname{tr}_2 S = \det \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix} + \det \begin{pmatrix} s_{11} & s_{13} \\ s_{13} & s_{33} \end{pmatrix} + \det \begin{pmatrix} s_{22} & s_{23} \\ s_{23} & s_{33} \end{pmatrix}.$$

Термин «след порядка k» для обозначения сумм главных миноров появился в конце XX века и впервые встречается в работах [3, 24].

Хорошо известно, что k-след обладает ортогональной инвариантностью, т. е. принимает одинаковые значения на матрицах S и $B^T S B$, где B — ортогональная матрица. Поэтому $\operatorname{tr}_k S = \sigma_k(\lambda)$, где $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ — набор собственных чисел симметричной матрицы S. Ввиду этого равенства в зарубежной математической литературе k-след часто определяют как элементарную симметрическую функцию от собственных чисел матрицы S (например, k-trace в [39]).

Таким образом, за счёт подстановки (2.3) неравенства Ньютона и Маклорена могут быть записаны как для коэффициентов гиперболического многочлена, так и для элементарных симметрических функций нескольких переменных или k-следов симметричной матрицы. Заметим, что чаще всего их записывают в терминах элементарных симметрических функций.

§ 3. Гиперболические многочлены нескольких переменных

Рассмотрим комплексный многочлен $P(x)$ степени m от нескольких переменных $x = (x_1, x_2, \ldots, x_n)$, и вектор $a \in \mathbb{R}^n$, $a \neq 0$. Допустим, многочлен $P(x)$ является однородным, т. е. $P(\lambda x) = \lambda^m P(x)$, $\lambda \in \mathbb{R}$. Согласно определению 1.2 многочлен $P(x)$ называется a-гиперболическим, или гиперболическим в направлении вектора a, если $P(a) \neq 0$ и многочлен $p(t) = P(at + x)$ имеет m вещественных корней при любом $x \in \mathbb{R}^n$. Поскольку $P(at + x) = P(a) \prod_{k=1}^{m} (t - t_k)$, $t_k \in \mathbb{R}$, то $P(x)/P(a)$ — многочлен с вещественными коэффициентами. Не умаляя общности, можно рассматривать вещественные гиперболические многочлены.

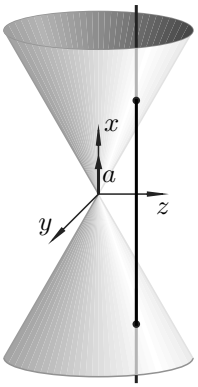


Рис. 1

С геометрической точки зрения, a-гиперболичность многочлена $P(x)$ значит, что любая прямая в пространстве $\mathbb{R}^n$ с направляющим вектором a пересекает гиперповерхность $P(x) = 0$ в m точках.

Приведём простейший пример. Однородный многочлен трёх переменных $P(x, y, z) = x^2 - y^2 - z^2$ является гиперболическим в направлении вектора $a = (1, 0, 0)$, поскольку любая прямая, параллельная оси x, пересекает поверхность $x^2 - y^2 - z^2 = 0$ ровно в двух точках (рис. 1).

Исходя из геометрического смысла можно сделать вывод, что если многочлен $P(x)$ является гиперболическим в направлении вектора a, то он является гиперболическим и в направлении любого вектора, пропорционального вектору a. В том числе многочлен остаётся гиперболическим, если поменять направление вектора a на противоположное. Но что будет, если незначительно изменить направление вектора a? Останется ли многочлен $P(x)$ гиперболическим в других направлениях, достаточно близких к вектору a?

Определение 3.1. Пусть $P(x)$ — a-гиперболический многочлен. Обозначим символом $K(P, a)$ наибольшее связное множество в пространстве $\mathbb{R}^n$, содержащее вектор a и состоящее из векторов, в направлении которых многочлен $P(x)$ является гиперболическим. Множество $K(P, a)$ называется конусом гиперболичности или конусом Гординга, соответствующим вектору a.

Конструктивные описания и геометрические свойства конуса Гординга даёт следующая теорема.

Теорема 3.2. *Пусть $P(x)$ — некоторый a-гиперболический многочлен. Тогда верны следующие утверждения о конусе $K(P, a)$:*

1) *конус $K(P, a)$ совпадает с той компонентой связности множества $\mathbb{R}^n \setminus \{x\colon P(x) = 0\}$, которая содержит вектор a;*
2) *конус $K(P, a)$ состоит ровно из таких векторов x, для которых многочлен $p(t) = P(at + x)$ имеет только отрицательные корни (это значит, что все его коэффициенты положительны в случае $P(a) > 0$);*
3) *конус $K(P, a)$ выпуклый;*
4) *функция $P^{1/m}(x)$ вогнута в конусе $K(P, a)$, где $P(a) > 0$.*

Эти ключевые результаты были впервые доказаны в работах Гординга [18, 19], а также передоказаны во множестве последующих работ разных авторов.

Пункт 2 теоремы 3.2 обосновывает следующее наблюдение. Несмотря на то, что a-гиперболический многочлен является и $(-a)$-гиперболическим, вектор $-a$ не попадает в компоненту связности $K(P,a)$, поскольку многочлен $P(at-a)=(t-1)^m P(a)$ имеет положительный корень. Значит, конус $K(P,a)$ является односторонним конусом с вершиной в нуле: если $x\in K(P,a)$, то $\lambda x\in K(P,a)$, $\lambda>0$. Конус $K(P,a)$ сопровождается конусом $K(P,-a)$. В общей ситуации множество всех векторов, в направлении которых многочлен является гиперболическим, распадается на чётное число конусов Гординга.

Из пункта 1 теоремы 3.2 вытекает, что каждый конкретный конус Гординга для многочлена $P(x)$ является компонентой положительности или компонентой отрицательности $P(x)$. Удобнее работать с компонентой положительности, поэтому далее рассматриваем $K(P,a)$, где $P(a)>0$.

Замечание 2. Рассмотрим разложение $P(at+x)$ по степеням t с коэффициентами, зависящими от a и от x:

$$P(at+x)=P(x)+P^a_{m-1}(x)t+P^a_{m-2}(x)t^2+\ldots+P^a_1(x)t^{m-1}+P(a)t^m. \quad (3.1)$$

Согласно пункту 2 теоремы 3.2 все коэффициенты этого многочлена положительны, если $x\in K(P,a)$, $P(a)>0$. Отсюда нетрудно вывести, что многочлен $P(x)$ возрастает в конусе $K(P,a)$, а именно:

$$P(x+y)>P(x), \quad x\in K(P,a),\ y\in \overline{K}(P,a). \quad (3.2)$$

Поэтому конусы Гординга называют ещё конусами монотонности.

Замечание 3. Если многочлен $P(x)$ является a-гиперболическим, то каждый из коэффициентов $P^a_k(x)$ в разложении (3.1) тоже является a-гиперболическим многочленом. Это следует из равенства

$$P^a_k(at+x)=\frac{1}{(m-k)!}\frac{d^{m-k}P(at+x)}{dt^{m-k}}$$

и того факта, что между корнями многочлена имеется корень его производной (теорема Ролля). При этом $K(P,a)\subset K(P^a_k,a)$.

§ 4. Примеры a-гиперболических многочленов и конусов Гординга

В этом параграфе исследуем конусы гиперболичности для конкретных однородных многочленов. Первые два примера обычно приводятся в статьях как тривиальные. Однако их разбор демонстрирует хорошие учебные

«задачи с параметром», для решения которых требуются только элементарные арифметические навыки и базовые знания из аналитической геометрии. Третий и четвёртый примеры наиболее важны для теории полностью нелинейных уравнений в частных производных.

Пример 1. Рассмотрим произвольный однородный многочлен первой степени:

$$P(x) = b_1x_1 + b_2x_2 + \ldots + b_nx_n, \quad b_k \in \mathbb{R}.$$

Гиперплоскость $P(x) = 0$ имеет ровно одно пересечение с любой прямой, направляемой вектором a, в том и только в том случае, когда вектор a не лежит в этой гиперплоскости. Поэтому многочлен $P(x)$ является гиперболическим в направлении любого такого вектора $a \in \mathbb{R}^n$, что $P(a) \neq 0$. Множество гиперболичности распадается на два конуса Гординга, на два полупространства.

Пример 2. Пусть имеется однородный многочлен второй степени:

$$P(x) = x_1^2 + x_2^2 + \ldots + x_k^2 - x_{k+1}^2 - x_{k+2}^2 - \ldots - x_n^2. \tag{4.1}$$

Рассмотрим два случая: $k > 1$ и $k = 1$. Проведём графический анализ числа корней уравнения $P(at + x) = 0$. Для этого перенесём последние $n - k$ квадратов в правую часть этого уравнения, раскроем скобки, выделим полные квадраты в обеих частях уравнения и введём следующие обозначения:

$$\begin{aligned}
&\alpha_1 = \sqrt{\sum_{i=1}^{k} a_i^2}, \quad \alpha_2 = \sqrt{\sum_{i=k+1}^{n} a_i^2}, \\
&\beta_1 = \frac{\sum_{i=1}^{k} a_i x_i}{\alpha_1}, \quad \beta_2 = \frac{\sum_{i=k+1}^{n} a_i x_i}{\alpha_2}, \\
&\gamma_1 = \frac{\sum_{i=1}^{k} x_i^2 \sum_{i=1}^{k} a_i^2 - \left(\sum_{i=1}^{k} a_i x_i\right)^2}{\sum_{i=1}^{k} a_i^2}, \\
&\gamma_2 = \frac{\sum_{i=k+1}^{n} x_i^2 \sum_{i=k+1}^{n} a_i^2 - \left(\sum_{i=k+1}^{n} a_i x_i\right)^2}{\sum_{i=k+1}^{n} a_i^2}.
\end{aligned} \tag{4.2}$$

Получим равенство:

$$(\alpha_1 t + \beta_1)^2 + \gamma_1 = (\alpha_2 t + \beta_2)^2 + \gamma_2. \tag{4.3}$$

Левая и правая части равенства (4.3) являются квадратичными функциями $y_1(t)$ и $y_2(t)$, графики которых — параболы. Ветви этих парабол направлены вверх, а вершина, по неравенству Коши — Буняковского, находится не ниже оси абсцисс. Предположим, что существует такой вектор a, в направлении

которого многочлен $P(x)$ гиперболический. Это значит, что данный вектор a обеспечивает параболам y_1 и y_2 наличие двух точек пересечения для любого вектора x. Не умаляя общности, можно считать, что вектор a обладает свойством:

$$\alpha_1^2 > \alpha_2^2,$$

т. е. скорость роста функции y_1 больше, чем скорость y_2.

Построим вектор $x = (x_p, x_q)$ так, чтобы вектор $x_p = (x_1, x_2, \ldots, x_k)$ был ортогонален вектору $a_p = (a_1, a_2, \ldots, a_k)$, а вектор $x_q = (x_{k+1}, x_{k+2}, \ldots, x_n)$ — вектору $a_q = (a_{k+1}, a_{k+2}, \ldots, a_n)$. Тогда

$$\beta_1 = \beta_2 = 0, \quad \gamma_1 = \sum_{i=1}^{k} x_i^2, \quad \gamma_2 = \sum_{i=k+1}^{n} x_i^2.$$

С геометрической точки зрения это означает, что вершины графиков функций y_1 и y_2 находятся друг над другом, причём абсцисса вершин равна нулю. Очевидно, что в случае $k > 1$ вектор x можно построить таким образом, чтобы параболы y_1 и y_2 не пересекались (рис. 2). Таким образом, пришли к противоречию с предположением об a-гиперболичности многочлена $P(x)$. Значит, не существует такого вектора a, в направлении которого многочлен $P(x)$ был бы гиперболическим.

Теперь рассмотрим многочлен (4.1) в случае $k = 1$. Получаем характеристический многочлен волнового оператора:

$$P(x) = x_1^2 - x_2^2 - x_3^2 - \ldots - x_n^2.$$

Уравнение $P(at + x) = 0$ можно привести к виду:

$$(a_1 t + x_1)^2 = (\alpha_2 t + \beta_2)^2 + \gamma_2. \tag{4.4}$$

Здесь параметры $\alpha_2, \beta_2, \gamma_2$ заданы формулами (4.2). Как и в равенстве (4.3), левая и правая части уравнения (4.4) представляют собой квадратичные

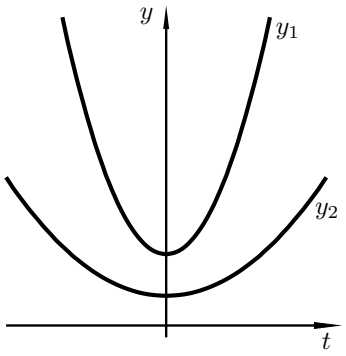


Рис. 2

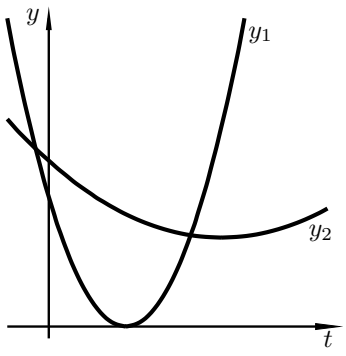


Рис. 3

функции $y_1(t)$ и $y_2(t)$, графики которых — параболы. Ветви этих парабол направлены вверх, однако вершина параболы y_1 расположена на оси абсцисс, а вершина y_2 находится не ниже этой оси. Несложно доказать, что для того чтобы параболы y_1 и y_2 имели две точки пересечения для любого вектора $x \in \mathbb{R}^n$, необходимо и достаточно потребовать, чтобы скорость роста функции y_1 была больше, чем скорость y_2 (рис. 3), или, что то же самое:

$$a_1^2 > \alpha_2^2 = \sum_{i=2}^{n} a_i^2.$$

Таким образом, при $k = 1$ многочлен (4.1) гиперболический в направлении любого вектора $a \in \mathbb{R}^n$, удовлетворяющего условию $P(a) > 0$. Множество таких векторов a образует две компоненты связности, т. е. два конуса Гординга.

Анализ, предпринятый здесь для многочлена (4.1), является на самом деле исчерпывающим для всех однородных многочленов второй степени, поскольку любая квадратичная форма может быть записана в каноническом виде (4.1) в подходящей системе координат. Заметим, что существуют и другие способы исследования многочлена (4.1) на гиперболичность.

Пример 3. Рассмотрим серию элементарных симметрических функций $\sigma_m(x)$, $x \in \mathbb{R}^n$, $1 \leqslant m \leqslant n$:

$$\sigma_m(x) = \sum_{i_1, i_2, \ldots, i_m} x_{i_1} x_{i_2} \ldots x_{i_m}.$$

Каждая функция $\sigma_m(x)$ является однородным многочленом степени m. Сначала обратимся к многочлену

$$\sigma_n(x) = x_1 x_2 \ldots x_n.$$

Он является гиперболическим в направлении любого такого вектора $a \in \mathbb{R}^n$, что $\sigma_n(a) \neq 0$. Действительно, многочлен

$$\sigma_n(at + x) = \prod_{k=1}^{n} (a_k t + x_k)$$

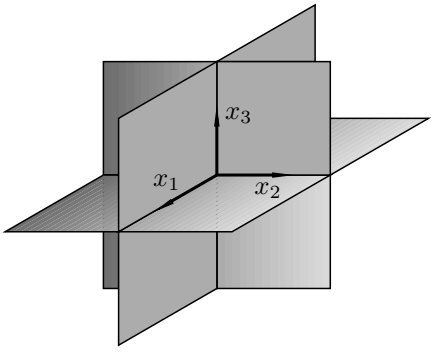


Рис. 4

имеет n вещественных корней $t = -x_k/a_k$ тогда и только тогда, когда все $a_k \neq 0$.

Множество гиперболичности этого многочлена распадается на 2^n конусов Гординга, ограниченных координатными плоскостями $x_k = 0$. В случае $n = 3$ имеем восемь координатных октантов (рис. 4).

Зафиксируем далее вектор $a = (1, 1, \ldots, 1)$, тогда

$$\sigma_n(at + x) = \sum_{m=0}^{n} \sigma_m(x) t^{n-m}.$$

Согласно замечанию 3 многочлены $\sigma_m(x)$, $m < n$, также являются гиперболическими в направлении данного вектора a. Из пункта 2 теоремы 3.2 и представления

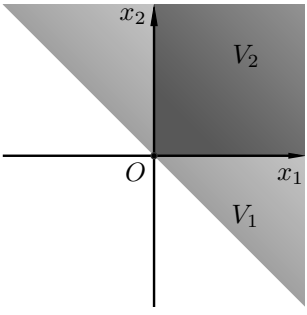
$$\sigma_m(at + x) = C_n^m \sum_{k=0}^{m} \frac{C_m^k}{C_n^k} \sigma_k(x) t^{m-k}$$

следует, что конус Гординга для многочлена $\sigma_m(x)$, соответствующий вектору $a = (1, 1, \ldots, 1)$, состоит из таких векторов x, для которых первые m симметрических функций положительны. Обозначим такие конусы следующим образом:

$$V_m = \{x \in \mathbb{R}^n \colon \sigma_k(x) > 0,\ k = 1, 2, \ldots, m\}. \tag{4.5}$$

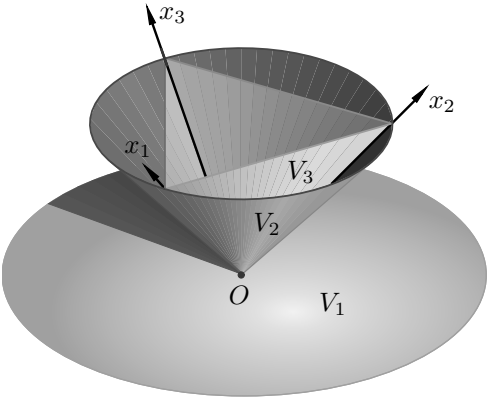
Очевидно, $V_1 \supset V_2 \supset \ldots \supset V_n$ и конус V_n состоит из всех векторов с положительными координатами. На рис. 5 изображены два таких конуса Гординга, соответствующих случаю $n = 2$ и многочленам $\sigma_1(x)$, $\sigma_2(x)$. На рис. 6 изображены три таких конуса Гординга, соответствующих случаю $n = 3$ и многочленам $\sigma_1(x)$, $\sigma_2(x)$, $\sigma_3(x)$. В этой ситуации конус V_1 — это полупространство, ограниченное плоскостью $x_1 + x_2 + x_3 = 0$, V_3 — первый координатный октант, V_2 — односторонний круговой конус, граница которого задана уравнением $x_1x_2 + x_2x_3 + x_1x_3 = 0$.

Понятно, что каждому конусу V_m соответствует парный конус, порождённый вектором $-a = (-1, -1, \ldots, -1)$. Как показано выше, для много-



$n = 2$

Рис. 5



$n = 3$

Рис. 6

члена $\sigma_n(x)$ имеются и другие пары конусов Гординга, в случаях $m = 1$, $m = 2$ других конусов нет. При остальных значениях m вопрос о наличии других конусов Гординга остаётся открытым.

Пример 4. Рассмотрим серию однородных многочленов, построенных на основе m-следа симметричной матрицы (см. определение 2.2). Каждый такой многочлен $\operatorname{tr}_m X$, $1 \leqslant m \leqslant n$, зависит от $n(n+1)/2$ переменных, организованных в виде матрицы $X = (x_{ij}) \in \operatorname{Sym}(n)$.

Сначала обратимся к многочлену $\operatorname{tr}_n X = \det X$. Очевидно, этот многочлен является гиперболическим в направлении единичной матрицы I, поскольку корнями уравнения $\det(It + X) = 0$ являются n собственных чисел симметричной матрицы X, взятых с обратным знаком. Согласно пункту 1 теоремы 3.2 вместе с единичной матрицей в конус Гординга этого многочлена попадают все положительно определённые матрицы. Аналогичные выводы верны и для матрицы $-I$. Таким образом, многочлен $\det X$ является гиперболическим в направлении любой положительно или отрицательно определённой матрицы A. Нетрудно показать, что эти два конуса Гординга исчерпывают всё множество гиперболичности данного многочлена, т. е. других конусов Гординга у многочлена нет.

Поскольку

$$\det(It + X) = \sum_{m=0}^{n} \operatorname{tr}_m X \cdot t^{n-m},$$

то согласно замечанию 3 многочлены $\operatorname{tr}_m X$, $m < n$, тоже являются I-гиперболическими, как и многочлен $\det X$. Из пункта 2 теоремы 3.2 и представления

$$\operatorname{tr}_m(It + X) = C_n^m \sum_{k=0}^{m} \frac{C_m^k}{C_n^k} \operatorname{tr}_k X \cdot t^{m-k}$$

следует, что конус Гординга для многочлена $\operatorname{tr}_m X$, соответствующий единичной матрице, состоит из симметричных матриц X, у которых все следы порядка от 1 до m положительны. Обозначим такой конус следующим образом:

$$K_m = \{X \in \operatorname{Sym}(n) \colon \operatorname{tr}_k X > 0,\ k = 1, 2, \ldots, m\}. \tag{4.6}$$

Определение 4.1. Матрица $X \in K_m$ называется m-положительной.

Конусы (4.6) впервые определила и исследовала Н. М. Ивочкина в работе [4]. Однако термин «m-положительная матрица» для наименования элементов из K_m введён совсем недавно тем же автором в работах [5, 27].

Остаётся открытым вопрос, имеет ли многочлен $\operatorname{tr}_m X$ в случае $2 < m < n$ другие конусы Гординга, кроме конуса m-положительных матриц и его «зеркальной» пары.

Для конусов (4.6) справедлива цепочка вложений $K_1 \supset K_2 \supset \ldots \supset K_n$. Причём K_n — это в точности конус положительно определённых матриц. Таким образом, понятие m-положительной матрицы является обобщением для положительно определённой матрицы.

Естественность такого обобщения подтверждается следующим. Для положительно определённых, или n-положительных, матриц широко известен критерий Сильвестра: матрица положительно определена тогда и только тогда, когда все её главные угловые миноры положительны. Совсем недавно на основе неравенств Ньютона (2.1) доказан аналог критерия Сильвестра для m-положительных матриц (см. статьи [6, 9]):

Теорема 4.2. *Пусть $X \in \mathrm{Sym}(n)$, $1 \leqslant m \leqslant n$, $1 \leqslant i_1, i_2, \ldots, i_{m-1} \leqslant n$ — произвольный набор различных номеров. Для того чтобы матрица X была m-положительной, необходимо и достаточно выполнение следующих условий*

$$\mathrm{tr}_m X > 0,\ \mathrm{tr}_{m-1} X^{\langle i_1 \rangle} > 0,\ \mathrm{tr}_{m-2} X^{\langle i_1, i_2 \rangle} > 0,\ \ldots,\ \mathrm{tr}_1 X^{\langle i_1, i_2, \ldots, i_{m-1} \rangle} > 0.$$

В теореме 4.2 символом $X^{\langle i_1, i_2, \ldots, i_k \rangle}$ обозначена матрица, полученная из матрицы X вычёркиванием k строк и k столбцов с номерами $i_1, i_2, \ldots, i_k$.

Нельзя не заметить очевидного сходства между примерами 3 и 4. Действительно, многочлены $\mathrm{tr}_m X$, заданные на матрицах $X \in \mathrm{Sym}(n)$, являются аналогами многочленов $\sigma_m(x)$, заданных на векторах $x \in \mathbb{R}^n$. Это обусловлено равенством $\mathrm{tr}_m X = \sigma_m(x)$, где $x = (x_1, x_2, \ldots, x_n)$ — набор собственных чисел матрицы X. Множества (4.5) и (4.6) являются конусами в пространствах $\mathbb{R}^n$ и $\mathrm{Sym}(n)$ соответственно. Векторный конус V_m — аналог матричного конуса K_m.

Описанные примеры a-гиперболических многочленов являются классическими. Другие примеры и методы построения a-гиперболических многочленов можно найти в работах [22, 23].

§ 5. Приложение конусов Гординга в FNPDE

В этом параграфе описана роль конусов Гординга в теории полностью нелинейных дифференциальных уравнений в частных производных (FNPDE) второго порядка.

5.1. Эллиптичность, конус допустимых функций

Все дифференциальные операторы в этом параграфе рассматриваются в стандартном пространстве $C^2(\Omega)$, состоящем из дважды непрерывно дифференцируемых функций $u(x)$, заданных в замкнутой ограниченной

области $\Omega \subset \mathbb{R}^n$. Символом u_x обозначен вектор градиента. Символом u_{xx} обозначена матрица Гессе, т. е. симметричная матрица, составленная из вторых частных производных функции u. Дифференциальный оператор второго порядка $F[u] = F(u_{xx}, u_x, u)$ называют линейным, если он представляет собой линейную комбинацию функции u и её частных производных:

$$F[u] = \sum_{i,j=1}^{n} a_{ij}(x) u_{x_i x_j}(x) + \sum_{k=1}^{n} a_k(x) u_{x_k}(x) + a(x)u(x).$$

Линейный оператор F называют эллиптическим, если матрица (a_{ij}) положительно определена в области Ω. Это свойство гарантирует разрешимость соответствующей задачи Дирихле.

В настоящее время теория линейных операторов вполне изучена и главное внимание уделяют примерам отсутствия линейности. Нелинейные дифференциальные операторы $F[u] = F(u_{xx}, u_x, u)$ различают по силе, глубине нелинейности. Чем слабее нелинейность, тем ближе оператор к линейному и тем проще приспособить к нему результаты линейной теории. Самым тяжёлым случаем считается нелинейность оператора F относительно вторых производных функции u. Такие операторы называют полностью нелинейными (fully nonlinear).

Теория FNPDE начала активно развиваться в 80-х годах XX века, и поначалу в ней использовались модификации привычных понятий линейной теории. В частности, на полностью нелинейный оператор F было распространено понятие эллиптичности. Рассмотрим функциональную зависимость оператора F от элементов матрицы u_{xx} и обозначим символом $F_{ij}[u]$ частную производную F по элементу $u_{x_i x_j}$:

$$F_{ij}[u] = \frac{\partial F[u]}{\partial u_{x_i x_j}}. \tag{5.1}$$

Заметим, что для линейного оператора $F_{ij}[u] = a_{ij}(x)$, а для полностью нелинейного производные F_{ij}, вообще говоря, зависят от функции u. Полностью нелинейный оператор F называют эллиптическим на функции u, если матрица $(F_{ij}[u])$ положительно определена в области Ω.

Большинство полностью нелинейных операторов не являются тотально эллиптическими, т. е. сохраняют эллиптичность только на некотором подмножестве $\mathbb{E}(\Omega)$ пространства $C^2(\Omega)$. Казалось бы естественно рассматривать задачу Дирихле для нелинейного оператора $F[u]$ в области его эллиптичности $\mathbb{E}(\Omega)$. Однако выяснилось, что $\mathbb{E}(\Omega)$ не является корректным множеством разрешимости. Основные причины заключаются в следующем.

Во-первых, для полностью нелинейных уравнений (в отличие от линейных) эллиптичность оператора не гарантирует единственность решения.

Простой пример наличия двух решений задачи Дирихле, на каждом из которых оператор $F[u]$ эллиптический, приведён в статье [27].

Во-вторых, на разрешимость краевой задачи для FNPDE радикальным образом влияют геометрические свойства границы области Ω, это принципиально не локальная задача, что предъявляет повышенные требования к выбору подходящих функций u.

Таким образом, сам термин «эллиптический оператор», унаследованный из линейной теории, оказывается недостаточно эффективным в теории полностью нелинейных уравнений. В исследовании FNPDE на первый план выходит построение специального конуса $\mathbb{K}(\Omega) \subset \mathbb{E}(\Omega) \subset C^2(\Omega)$, обеспечивающего оператору F не только эллиптичность, но и согласование с геометрическими характеристиками $\partial\Omega$, и ряд других полезных свойств, необходимых для доказательства теорем разрешимости.

В фундаментальной работе [14] 1985 года, посвящённой Л. Гордингу, впервые замечено, что конусы $\mathbb{K}(\Omega)$ можно строить по образцу конусов Гординга. При этом функции из конуса $\mathbb{K}(\Omega)$ названы допустимыми. Опишем в общих чертах идею этого построения.

Пусть имеется однородный многочлен $P(X)$ степени m, заданный на матрицах $X = (x_{ij}) \in \mathrm{Sym}(n)$. Предположим, многочлен $P(X)$ является I-гиперболическим, а соответствующий конус Гординга $K = K(P, I)$ является компонентой его положительности и при этом содержит все положительно определённые матрицы. Сопоставим многочлену оператор

$$F[u] = P(u_{xx}).$$

Назовём функцию $u \in C^2(\Omega)$ допустимой для этого оператора, если её матрица Гессе принадлежит конусу K в каждой точке области Ω. Обозначим конус допустимых функций символом $\mathbb{K}(\Omega)$.

Докажем эллиптичность нормированной версии оператора F в конусе $\mathbb{K}(\Omega)$. Введём обозначение

$$\mathcal{F}[u] = F^{1/m}[u].$$

Рассмотрим неотрицательно определённую матрицу $\xi \times \xi \in \overline{K}(P, I)$, $\xi \in \mathbb{R}^n$. Используем монотонность (замечание 2) и вогнутость функции $P^{1/m}(X)$ в конусе Гординга (теорема 3.2) для вывода цепочки соотношений:

$$\mathcal{F}[u] < \mathcal{F}[u + \xi \times \xi] \leqslant \mathcal{F}[u] + \sum_{i,j=1}^{n} \mathcal{F}_{ij}[u]\xi_i\xi_j, \quad u \in \mathbb{K}(\Omega).$$

Отсюда следует, что матрица $(\mathcal{F}_{ij}[u])$ положительно определена, а поскольку

$$F_{ij}[u] = m\mathcal{F}^{m-1}[u]\mathcal{F}_{ij}[u],$$

матрица $(F_{ij}[u])$ тоже положительно определена. Оператор $F[u]$ эллиптический на любой допустимой функции u, что доказывает включение $\mathbb{K}(\Omega) \subset \mathbb{E}(\Omega)$ (обратное включение может быть неверно).

Далее продолжаем демонстрировать естественность выбора $\mathbb{K}(\Omega)$ на примере конкретного оператора, порождённого I-гиперболическим многочленом.

5.2. Конус m-допустимых функций для m-гессиановского оператора

В § 4 (пример 4) описана серия I-гиперболических многочленов $\operatorname{tr}_m X$ (m-след матрицы X), $1 \leqslant m \leqslant n$. Операторы

$$F_m[u] = \operatorname{tr}_m(u_{xx}) \tag{5.2}$$

называются m-гессиановскими, и так же называются соответствующие уравнения (в зарубежной литературе чаще встречаются под названием k-hessian equation). Класс операторов F_m соединяет классический линейный оператор с классическим представителем нелинейной теории. При $m = 1$ имеем оператор Лапласа:

$$F_1[u] = \operatorname{tr}_1(u_{xx}) = \Delta u.$$

При $m = 2, 3, \ldots, n$ операторы F_m полностью нелинейные, причём в случае $m = n$ получаем оператор Монжа — Ампера:

$$F_n[u] = \operatorname{tr}_n(u_{xx}) = \det u_{xx}.$$

В программной статье [14] L. Caffarelly, L. Nirenberg и J. Spruck рассматривали m-гессиановские уравнения как эталонный пример общей теории FNPDE. Примерно с 1980-х годов m-гессиановскими уравнениями занимаются такие зарубежные учёные, как N. S. Trudinger, X.-J. Wang [28, 39], K. S. Chou, B. Guan. В российской математике в этом направлении работает только Н. М. Ивочкина и её ученики (см. серию работ Н. М. Ивочкиной, Н. В. Филимоненковой).

Рассмотрим задачу Дирихле для простейшего m-гессиановского уравнения с положительной правой частью:

$$F_m[u] = f > 0, \quad u|_{\partial\Omega} = \varphi. \tag{5.3}$$

Очевидно, что решение задачи (5.3) для 1-гессиановского уравнения (уравнения Пуассона) надо искать среди функций u, у которых матрица Гессе обладает в области Ω положительным 1-следом: $\operatorname{tr}_1(u_{xx}) = \triangle u > 0$. Как известно, задачу (5.3) для n-гессиановского уравнения (уравнения Монжа —

Ампера) целесообразно рассматривать на множестве строго выпуклых функций. Естественное множество разрешимости для задачи (5.3) при $1 < m < n$ впервые было построено Н. М. Ивочкиной в работе [4] 1983 года, причём ещё до её знакомства с теорией Гординга. Впоследствии оказалось, что построенный ею конус допустимых функций для оператора F_m есть не что иное, как функциональный аналог матричного конуса Гординга для многочлена $\mathrm{tr}_m X$. Напомним (см. § 4), что конус Гординга для многочлена $\mathrm{tr}_m X$, содержащий единичную матрицу, состоит из m-положительных матриц и обозначен символом K_m. При этом $K_1 \supset K_2 \supset \ldots \supset K_n$, где K_n — это конус всех положительно определённых матриц.

Определение 5.1. Функция u называется m-допустимой в области Ω, если матрица u_{xx} является m-положительной во всей этой области.

Конус m-положительных матриц (4.6) порождает конус m-допустимых функций:

$$\mathbb{K}_m(\Omega) = \{u \in C^2(\Omega) \colon F_k[u](x) > 0,\ k = 1, 2, \ldots, m,\ x \in \Omega\}.$$

Очевидна цепочка включений $\mathbb{K}_1(\Omega) \supset \mathbb{K}_2(\Omega) \supset \ldots \supset \mathbb{K}_n(\Omega)$, причём любая n-допустимая функция является строго выпуклой в традиционном смысле.

Опишем коротко идею исследования классической разрешимости задачи (5.3). В матричном конусе K_m известно много полезных алгебраических соотношений, список которых пополняется до сих пор. В частности, фирменной чертой K_m является справедливость неравенств Маклорена (2.2), записанных для k-следов с помощью подстановки (2.3). Это позволяет усилить эллиптичность оператора F_m до равномерной эллиптичности: оператор F_m является равномерно эллиптическим на каждом m-допустимом решении уравнения (5.3) при условии $f \geqslant \nu > 0$ в Ω и при наличии априорной оценки величины $\|u\|_{C^2(\Omega)}$. Кроме того, благодаря всё тем же неравенствам Маклорена и вогнутости функции $P^{1/m}(X)$ в конусе K_m, для уравнения (5.3) имеется аналог принципа максимума Александрова, который «вытесняет» C^2-оценки на границу области Ω. Для получения оценок на границе используются специальные барьеры. Их построение требует наличия у поверхности $\partial\Omega$ особых геометрических свойств, о которых будет сказано ниже. Затем по известным методикам, разработанным для равномерно эллиптических уравнений, априорную C^2-оценку можно поднять до такого класса гладкости, который соответствует гладкости данных задачи (5.3). Всё это в конечном счёте позволяет использовать метод непрерывности (продолжения по параметру) для доказательства следующей теоремы разрешимости:

Теорема 5.2. *Пусть* $\varphi \in C^{l+\alpha}(\partial\Omega)$, $f \in C^{l-2+\alpha}(\Omega)$, $\partial\Omega \in C^{l+\alpha}$, $f \geqslant \nu > 0$ *в* Ω, $l \geqslant 4$, $0 < \alpha < 1$. *Пусть, кроме того,*

$$\partial\Omega \text{ — строго } (m-1)\text{-выпуклая поверхность.} \tag{5.4}$$

Тогда существует единственное m*-допустимое решение* $u \in C^{l+\alpha}(\Omega)$ *задачи* (5.3).

Здесь символ $C^{l+\alpha}(\Omega)$ обозначает пространство функций, у которых производные порядка l непрерывны по Гёльдеру с показателем α. В частности, при $\alpha = 0$ пространство $C^{l+\alpha}(\Omega)$ состоит из l раз непрерывно дифференцируемых функций.

Наиболее полное доказательство теоремы 5.2 с построением априорных оценок в пространстве $C^2(\Omega)$ изложено в статье [16].

В теореме 5.2 первостепенную роль играет требование строгой $(m-1)$-выпуклости для поверхности $\partial\Omega$. Оно описывается при помощи сравнительно новой конструкции в дифференциальной геометрии, а именно: при помощи матрицы кривизны $\mathcal{K}[\Gamma]$, которая введена в работах Н. М. Ивочкиной как новый геометрический инвариант для произвольной C^2-гладкой гиперповерхности Γ. Не вдаваясь в детали определения матрицы кривизны, отметим только, что это симметричная матрица размера $(n-1) \times (n-1)$ и что это вполне вычислимая характеристика наподобие первой и второй квадратичных форм поверхности. Матрица $\mathcal{K}[\Gamma]$ может быть построена в явном виде в типичных случаях (см. [5, 6]).

Определение 5.3. Значения $\operatorname{tr}_k \mathcal{K}[\Gamma]$ (k-следы матрицы кривизны) называются k-кривизнами гиперповерхности Γ, $1 < k < n-1$. Гиперповерхность Γ называется строго p-выпуклой, если её матрица кривизны является в каждой точке p-положительной, т. е. $\mathcal{K}[\Gamma] \in K_p$.

Поскольку конус $K_{n-1} \subset \operatorname{Sym}(n-1)$, состоящий из $(n-1)$-положительных матриц, в точности совпадает с конусом положительно определённых матриц, $(n-1)$-выпуклая гиперповерхность является строго выпуклой в классическом смысле.

Известно, что собственные числа матрицы $\mathcal{K}[\Gamma]$ совпадают с главными кривизнами гиперповерхности Γ. Следовательно, k-кривизны есть не что иное, как элементарные симметрические функции от главных кривизн. В частности, $(n-1)$-кривизна — это гауссова кривизна гиперповерхности.

Для проверки p-выпуклости произвольной гиперповерхности Γ нужно удостовериться в положительности её k-кривизн, начиная с 1-кривизны и до p-кривизны включительно. Однако для замкнутой гиперповерхности достаточно, чтобы её p-кривизна была положительной. Поэтому усло-

вие (5.4) сводится к положительности $(m-1)$-кривизны $\partial\Omega$:

$$\operatorname{tr}_{m-1} \mathcal{K}[\partial\Omega] > 0. \tag{5.5}$$

При $m = n$ теорема 5.2 утверждает известный факт: разрешимость задачи Дирихле для уравнения Монжа — Ампера в области Ω, ограниченной гиперповерхностью с положительной гауссовой кривизной.

Естественность условия (5.4)–(5.5) при $m < n$ подтверждается тем, что поверхность уровня m-допустимой функции является строго $(m-1)$-выпуклой. Отсюда очевидно, что требование (5.4)–(5.5) необходимо для существования m-допустимого решения задачи Дирихле (5.3) с постоянным граничным условием. На самом деле в этой ситуации справедлива ещё более сильная «теорема несуществования» (см. [6]): даже если f и $\partial\Omega$ достаточно гладкие, но при этом $(m-1)$-кривизна поверхности $\partial\Omega$ вырождается хотя бы в одной точке, то задача (5.3) не имеет не только m-допустимых решений, но вообще не имеет C^2-гладких решений.

Всё это говорит о том, что исследовать задачу Дирихле для m-гессиановского уравнения целесообразно именно в конусе m-допустимых функций.

Замечание 4. В зарубежных публикациях по FNPDE, начиная с работы [14], закрепилась традиция определять m-гессиановские операторы и уравнения при помощи элементарных симметрических функций:

$$F_m[u] = \sigma_m(\lambda(u_{xx})), \tag{5.6}$$

где $\lambda(u_{xx}) = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ — набор собственных чисел матрицы Гессе. В этом случае теория строится на основе a-гиперболических многочленов $\sigma_m(x)$, $x \in \mathbb{R}^n$, $a = (1, 1, \ldots, 1)$, и соответствующих конусов Гординга V_m (см. § 4, пример 3). Такой подход позволяет иметь дело с векторами вместо матриц. Это упрощает внешний вид некоторых формул и условий. Например, условие (5.4)–(5.5) преобразуется к виду:

$$\sigma_{m-1}(\kappa_1, \kappa_2, \ldots, \kappa_n) > 0,$$

где $\kappa_1, \kappa_2, \ldots, \kappa_n$ — главные кривизны $\partial\Omega$. Однако техническая проверка данного условия и вообще любые конкретные вычисления при этом сильно затрудняются. Например, возникают проблемы с дифференцированием оператора $F_m[u]$ в области Ω. Поэтому форму записи (5.6) следует считать оправданной только исторически.

На примере этого краткого обзора m-гессиановских уравнений видно, каким образом конусы Гординга генерируют новые системы алгебро-геометрических понятий: распределение поверхностей по классам p-выпуклости

согласовано с распределением функций по конусам m-допустимости, и оба порождены распределением симметричных матриц по конусам m-положительности. Подробное описание этих конструкций, основанных на теории Гординга, можно найти в работах [5, 6, 16, 17, 27].

В заключение отметим, что уравнение (5.3) является простейшим стационарным уравнением, порождённым I-гиперболическим многочленом $\mathrm{tr}_m X$. Помимо него изучаются эволюционные m-гессиановские уравнения, уравнения p-кривизны, уравнения, порождённые отношениями многочленов $\mathrm{tr}_m X$ разной степени, и т. д.

Список литературы

[1] *Арнольд В. И.* Гиперболические многочлены и отображения Вандермонда // Функц. анализ и его прил. 1986. Т. 20, вып. 2. С. 52–53.

[2] *Ивочкина Н. М.* Интегральный метод барьерных функций и задача Дирихле для уравнений с операторами типа Монжа — Ампера // Матем. сб. 1980. Т. 112(154), № 2(6). С. 193–206.

[3] *Ивочкина Н. М.* Мини-обзор основных понятий в теории полностью нелинейных дифференциальных уравнений второго порядка (англ.) // Зап. научн. сем. ПОМИ. 1997. Т. 249. С. 199–211.

[4] *Ивочкина Н. М.* Описание конусов устойчивости, порождаемых дифференциальными операторами типа Монжа — Ампера // Матем. сб. 1983. Т. 122(164), № 2(10). С. 265–275.

[5] *Ивочкина Н. М.* От конусов Гординга к p-выпуклым гиперповерхностям // Соврем. матем. Фунд. напр. 2012. Т. 45. С. 94–104.

[6] *Ивочкина Н. М., Филимоненкова Н. В.* О новых структурах в теории полностью нелинейных уравнений // Соврем. матем. Фунд. напр. 2015. Т. 58. С. 82–95.

[7] *Ньютон И.* Всеобщая арифметика, или Книга об арифметическом синтезе и анализе. М.: Изд. АН СССР, 1948.

[8] *Петровский И. Г.* О диффузии волн и лакунах для гиперболических уравнений (англ.) // Матем. сб. 1945. Т. 17(59), № 3. С. 289–370.

[9] *Филимоненкова Н. В.* Критерий Сильвестра для m-положительных матриц // Препринт Санкт-Петербургского математического общества. 2014. № 7.

[10] *Харди Г. Г., Литтлвуд Дж. Е., Полиа Г.* Неравенства. М.: ИЛ, 1948.

[11] *Atiyah M. F., Bott R., Gårding L.* Lacunas for hyperbolic differential operators with constant coefficients. I // Acta Math. 1970. V. 124. P. 109–189.

[12] *Bauschke H. H., Güler O., Lewis A. S., Sendov H. S.* Hyperbolic polynomials and convex analysis // Canad. J. Math. 2001. V. 53(3). P. 470–488.

[13] *Brändén P.* Obstructions to determinantal representability // Adv. Math. 2011. V. 226, № 2. P. 1202–1212.

[14] *Caffarelly L., Nirenberg L., Spruck J.* The Dirichlet problem for nonlinear second order elliptic equations III. Functions of the eigenvalues of the Hessian // Acta Math. 1985. V. 155. P. 261–301.

[15] *Chaillou J.* Hyperbolic differential polynomials and their singular perturbations. Dordrecht — Boston — London: D. Reidel Publishing Company, 1979. (Mathematics and its applications; V. 3).

[16] *Filimonenkova N. V.* On the classical solvability of the Dirichlet problem for nondegenerate m-Hessian equations // J. Math. Sci. 2011. V. 178, № 6. P. 666–694.

[17] *Filimonenkova N. V., Ivochkina N. M.* On algebraic and geometric conditions in the theory of Hessian equations // J. Fixed Point Theory and Applications. 2014. V. 16, № 1–2. P. 11–25.

[18] *Gårding L.* An inequality for hyperbolic polynomials // J. Math. Mech. 1959. V. 8. P. 957–965.

[19] *Gårding L.* Linear hyperbolic partial differential equations with constant coefficients // Acta Math. 1951. V. 85. P. 1–62.

[20] *Güler O.* Hyperbolic polynomials and interior point methods for convex programming // Math. Oper. Res. 1997. V. 22, № 2. P. 350–377.

[21] *Gurvits L.* Combinatorial and algorithmic aspects of hyperbolic polynomials. Electronic Colloquium on Computational Complexity, Report № 70, 2004.

[22] *Harvey F. R., Lawson H. B., Jr.* Gårding's theory of hyperbolic polynomials // Comm. Pure Appl. Math. 2013. V. 66, № 7. P. 1102–1128.

[23] *Hörmander L.* Linear partial differential operators. Springer-Verlag, 1963.

[24] *Ivochkina N. M.* On the Dirichlet problem for fully nonlinear parabolic equations // Зап. научн. сем. ПОМИ. 1996. Т. 233 С. 101–111.

[25] *Ivochkina N. M.* Weakly first-order interior estimates and Hessian equations // Зап. научн. сем. ПОМИ. 2006. Т. 336 С. 55–66.

[26] *Ivochkina N. M., Filimonenkova N. V.* On the backgrounds of the theory of m-Hessian equations // Comm. Pure and Appl. Anal. 2013. V. 12, № 4. P. 1687–1703.

[27] *Ivochkina N. M., Prokof'eva S. I. Yakunina G. V.* The Gårding cones in the modern theory of fully nonlinear second order differential equations // J. Math. Sci. 2012. V. 184, № 3. P. 295–315.

[28] *Ivochkina N. M., Trudinger N., Wang X.-J.* The Dirichlet problem for degenerate Hessian equations // Comm. Partial Diff. Equations. 2004. V. 29. P. 219–235.

[29] *Kostov V.* Topics on hyperbolic polynomials in one variable. Paris: Société Mathématique de France, 2011. (Panoramas et Synthèses; V. 33).

[30] *Kummer M., Plaumann D., Vinzant C.* Hyperbolic polynomials, interlacers and sums of squares. Preprint, 2013. arXiv:1212.6696.

[31] *Lax A.* On Cauchy's problem for partial differential equations with multiple characteristics // Comm. Pure Appl. Math. 1956. V. 9. P. 135–169.

[32] *Maclaurin C.* A second letter to Martin Folkes, Esq: concerning the roots of equations, with the demonstration of other rules in algebra // Phil. Transactions. 1729. V. 36. P. 59–96.

[33] *Newton I.* Arithmetica universalis; sive De compositione et resolutione arithmetica liber. 1707.

[34] *Nuij W.* A note on hyperbolic polynomials // Math. Scand. 1968. V. 23. P. 69–72.

[35] *Pemantle R.* Hyperbolicity and stable polynomials in combinatorics and probability // Current Developments in Mathematics, 2011. Somerville, MA: Int. Press, 2012. P. 57–123.

[36] *Rellich F.* Perturbation theory of eigenvalue problems. New York: Gordon and Breach Sci. Publ., 1969.

[37] *Renegar J.* Hyperbolic programs, and their derivative relaxations // Found. Comput. Math. 2006. V. 6, № 1. P. 59–79.

[38] *Svensson S. L.* Necessary and sufficient conditions for the hyperbolicity of polynomials with hyperbolic principal part // Ark. fur Mat. 1969. V. 8. P. 145–162.

[39] *Wang X.-J.* The k-Hessian equation. Springer-Verlag, 2009. (Lecture Notes in Math.; V. 1977). P. 177–252.

Н. В. Филимоненкова, Санкт-Петербургский политехнический университет Петра Великого

nf33@yandex.ru

П. А. Бакусов, Санкт-Петербургский государственный архитектурно-строительный университет

bakusovpavel@gmail.com

Экспоненциальные диофантовы уравнения и сумма цифр числа

В. М. Журавлёв, П. И. Самовол

> Если я смогу расширить в этом пункте границы Арифметики, то Вы даже не поверите, какие чудесные предложения мы сможем отсюда получить.
>
> *Из письма Ферма к Мерсенну*

> Число «66» ничем не успело зарекомендовать себя в глазах Стёпы. Оно состояло в визуальном родстве с нулём, что делало его не до конца понятным. Кроме того, «66» приходилось родственником другому известному числу, «666». Иного это напугало бы, но Стёпа гораздо сильнее опасался числа «661», которое делилось на «43» без остатка.
>
> *Виктор Пелевин.* Числа

§ 1. Введение

Однажды французский композитор и теоретик музыки Филипп де Витри заказал Льву Герсониду небольшое математическое исследование. Филиппа де Витри интересовали гармонические[1)] числа, которые он использовал при написании музыкальных произведений — мотетов. В 1343 году Лев Герсонид завершил исследование и написал короткое сочинение, которое сохранилось под названием «De Numeris Harmonicis». В этой работе было доказано, что существуют только четыре пары гармонических чисел, отличающихся на единицу — это пары $(1;2)$, $(2;3)$, $(3;4)$ и $(8;9)$. На современном математическом языке можно сказать, что Герсонид решил

[1)] Числа вида 2^a3^b.

в целых неотрицательных числах уравнение

$$2^a 3^b + 1 = 2^c 3^d.$$

Если использовать свойства делимости на 2 и на 3, то легко видеть, что решение этого уравнения сводится к двум следующим случаям.

Задача 1. Найти все решения уравнений в целых неотрицательных числах n и m:

$$\text{а) } 1 + 2^n = 3^m; \quad \text{б) } 1 + 3^m = 2^n. \tag{1}$$

В такой простой формулировке задача по решению этих уравнений давно стала фольклорной, и мы можем найти её в разных источниках [3, 9, 10]. Впрочем, для решения этой задачи школьнику нужно не только уметь правильно делить с остатком (в отличие от Стёпы из эпиграфа[2)]), но также знать основы теории делимости и иметь некоторые навыки решения задач с помощью сравнений по модулю. К слову сказать, современное решение этих уравнений, доступное для понимания школьника, занимает чуть больше двух абзацев.

В современной теории чисел уравнения, решения которых ищутся в натуральных, целых или рациональных числах, называются диофантовыми уравнениями. Из дошедших до нашего времени источников можно узнать, что сам Диофант умел решать не только линейные и квадратные уравнения, но также некоторые уравнения третьей, четвёртой и даже шестой степени. Если неизвестные величины находятся в показателях степеней членов диофантового уравнения, то такие уравнения называются экспоненциальными или показательными диофантовыми уравнениями. Уравнения такого вида гораздо труднее поддаются изучению в отличие от полиномиальных диофантовых уравнений. Таким образом, с большой долей уверенности можно сказать, что Герсонид первым начал исследование экспоненциальных диофантовых уравнений.

§ 2. Некоторые обобщения

Вышеуказанные уравнения являются частными случаями диофантового уравнения

$$a^x - b^y = k, \tag{2}$$

где a и b — натуральные параметры, $k \neq 0$ — целое число.

Поисками решений этого диофантового уравнения математики активно занимались в середине прошлого века и доказали, что количество его

[2)] Элементарная проверка показывает, что $661 = 43 \cdot 15 + 16 \equiv 16 \pmod{43}$, т. е. 661 нацело на 43 не делится.

решений конечно (см., например, [13, с. 51]). Более того, если $k = 1$, то это уравнение имеет не более одного решения за исключением случая $a = 3$ и $b = 2$, в котором имеется ровно два решения, указанных Герсонидом.

На рубеже веков удалось решить ещё одно из известных обобщений уравнения (1) — *уравнение Каталана*

$$x^u - y^v = 1, \quad \text{где } u, v > 1.$$

Подробно об этом уравнении можно прочитать в статье [8].

В рамках этой статьи мы хотим провести небольшое математическое исследование. Будем обобщать уравнения из задачи 1, увеличивая количество слагаемых в левой части. Мы решим несколько таких уравнений и по ходу исследования попытаемся ответить на возникающие вопросы.

Для начала нас будет интересовать, в каких случаях степень тройки можно представить в виде суммы единицы и двух различных степеней двойки?

Задача 2. Найти все решения уравнения в натуральных числах a, b, m

$$1 + 2^a + 2^b = 3^m, \quad \text{где } a < b. \tag{3}$$

Отметим, что в условии задачи мы оставляем строгое неравенство $a < b$, поскольку при $a = b$ мы просто попадаем в случай задачи 1. Затем попробуем узнать, в каких случаях степень тройки можно представить в виде суммы единицы и трёх различных степеней двойки.

Задача 3. Найти все решения уравнения в натуральных числах

$$1 + 2^a + 2^b + 2^c = 3^m, \quad \text{где } a < b < c. \tag{4}$$

Общий случай с k слагаемыми подразумевает серьёзное знание темы. Поэтому попытаемся его конкретизировать.

Диофантовы уравнения могут иметь и конечное число решений (в частности, ни одного), и бесконечное. Исходя из этого сформулируем следующую задачу.

Задача 4. Зафиксируем натуральное число k. Конечно или бесконечно количество решений в натуральных числах уравнения

$$1 + 2^{a_1} + 2^{a_2} + 2^{a_3} + \ldots + 2^{a_k} = 3^m, \quad \text{где } a_1 < a_2 < \ldots < a_k? \tag{5}$$

Этот общий вопрос на сегодняшний день не имеет простого решения. Он связан с очень серьёзными фундаментальными теоремами, и его подробное обсуждение выходит за рамки этой небольшой заметки. Тем не менее, мы попытаемся поработать с большим числом слагаемых, чтобы определить возможные подходы к решению этой задачи элементарными методами.

§ 3. Несколько частных случаев

Любое исследование начинается с рассмотрения частных случаев. Мы не будем отступать от этого принципа и поэтому приведём решение п. а) первой задачи. Итак, решаем уравнение $1 + 2^n = 3^m$. Как мы и обещали, попытаемся втиснуть решение в два абзаца, сконцентрировавшись на математической идее и опуская излишние подробности.

Решение задачи 1а). Если m нечётно, т. е. $m = 2q + 1$, то $3^m \equiv 3^{2q+1} \equiv 3 \pmod 8$, откуда $1 + 2^n \equiv 3 \pmod 8$. Последнее сравнение выполняется только для $n = 1$, тогда $m = 1$.

Если m чётно, то $m = 2q$. Тогда $2^n = (3^q - 1)(3^q + 1)$. Последнее равенство возможно, только если каждый сомножитель в правой части является степенью двойки. Имеем для некоторых натуральных b, c

$$\begin{cases} 3^q - 1 = 2^b, \\ 3^q + 1 = 2^c \end{cases} \quad \Rightarrow \quad 2 = 2^c - 2^b \quad \Rightarrow \quad c = 2,\ b = 1,\ q = 1.$$

Откуда $m = 2$, $n = 3$.

Итак, ответ для задачи 1а) состоит из двух пар: $(n; m) \in \{(1; 1), (3; 2)\}$.

Надеемся, что подобным образом читатели решат уравнение $1 + 3^m = 2^n$ и подтвердят результат, полученный Герсонидом.

Упражнение 1. Найти все решения уравнений в натуральных числах m, n и $q > 7$:

$$\text{а) } 1 + 2^n = 5^m; \quad \text{б) } 1 + 2^n = 7^m; \quad \text{в) } 1 + 2^n = q^m;$$
$$\text{г) } 2^n - 1 = 5^m; \quad \text{д) } 2^n - 1 = 7^m; \quad \text{е) } 2^n - 1 = q^m.$$

Перейдём теперь к уравнению $1 + 2^a + 2^b = 3^m$, где $1 \leqslant a < b$.

Решение задачи 2 (I способ). Правая часть уравнения делится на три, следовательно, и левая часть делится на три. Это возможно только в том случае, когда оба показателя степеней двоек — чётные числа. Значит, $a = 2a_1$ и $b = 2b_1$. Тогда имеем $4^{a_1} + 4^{b_1} = 3^m - 1$. Поскольку левая часть полученного уравнения делится на 4, то и правая его часть должна делиться на 4. Поэтому m чётно, т. е. $m = 2q$. Разложив на множители, получим

$$2^{2a_1} \cdot (1 + 4^{b_1 - a_1}) = (3^q - 1)(3^q + 1).$$

Числа $3^q - 1$ и $3^q + 1$ — последовательные чётные числа. Поэтому одно из них делится на 2 только в первой степени. Рассмотрим два возможных случая.

В первом случае имеем

$$\begin{cases} 3^q - 1 = 2^{2a_1-1} \cdot t, \\ 3^q + 1 = 2^{2a_1-1} \cdot t + 2, \end{cases} \quad \text{где } t \text{ — нечётное число.}$$

Тогда

$$2^{2a_1} \cdot (1 + 4^{b_1-a_1}) = 2^{2a_1} t(2^{2a_1-2}t + 1).$$

Разделив обе части уравнения на 2^{2a_1}, получим

$$1 + 4^{b_1-a_1} = t(2^{2a_1-2}t + 1).$$

Далее,

$$t - 1 = 4^{b_1-a_1} - 2^{2a_1-2}t^2 = (2^{b_1-a_1} - 2^{a_1-1}t)(2^{b_1-a_1} + 2^{a_1-1}t).$$

Если нечётное число $t \geqslant 3$, то левая часть уравнения положительна, следовательно, и правая положительна, и мы имеем $2^{b_1-a_1} - 2^{a_1-1}t \geqslant 1$. Но тогда $t - 1 \geqslant 1 \cdot (1 + t)$. Получаем противоречие. Остаётся $t = 1$.

При $t = 1$ мы получаем уравнение $3^q - 1 = 2^{2a_1-1}$, которое мы решали выше. Для $q = 1$, $m = 2$ получим противоречие с неравенством $1 \leqslant a < b$. Значит, $q = 2$ и $m = 4$. Тогда $a_1 = 2$, $a = 4$, откуда $b = 6$.

Во втором случае мы должны записать

$$\begin{cases} 3^q - 1 = 2^{2a_1-1} \cdot t - 2, \\ 3^q + 1 = 2^{2a_1-1} \cdot t, \end{cases} \quad \text{где } t \text{ — нечётное число.}$$

Аналогично получаем $1 + 4^{b_1-a_1} = t(2^{2a_1-2}t - 1)$ и далее

$$t + 1 = 2^{2a_1-2} \cdot t^2 - 4^{b_1-a_1} = (2^{a_1-1}t - 2^{b_1-a_1})(2^{a_1-1}t + 2^{b_1-a_1}).$$

Последнее равенство для нечётных значений t возможно при $a < b$ только в случае $a_1 = 1$. Но тогда $t + 1 \geqslant 1 \cdot (t + 2)$. Получаем противоречие.

В результате мы нашли единственное решение диофантова уравнения: $(a; b; m) \in \{(4; 6; 4)\}$.

Упражнение 2 (LXI ММО, 1998, задача 11.4, [11]). Решите в натуральных числах уравнение $3^x + 4^y = 5^z$.

Упражнение 3 (47 IMO, задача 4, Словения 2006). Найдите все пары (x, y) целых чисел такие, что $1 + 2^x + 2^{2x+1} = y^2$.

Решение задачи 2 не кажется слишком сложным и экстраординарным. Мы использовали те же методы, что и при решении задачи 1, а именно

сравнения по модулю и разложение на множители. Кроме того, на одном из этапов решения задачи 2 нам удалось свести её к задаче 1, что наталкивает на мысль о возможности применения метода математической индукции. Именно поэтому вначале у авторов возникло ложное чувство, что трудности, возникающие при дальнейшем увеличении числа слагаемых, будут преодолимыми. Однако уже при попытках рассмотреть следующий случай, а именно представление степени тройки в виде суммы единицы и трёх различных степеней двойки, обнаружилось много сюрпризов.

Авторы предлагают пытливым читателям хотя бы на пару часов отложить чтение этой статьи и попробовать самостоятельно решить задачу 3, а затем сравнить результаты.

Итак, продолжим.

Нам понадобятся некоторые классические результаты, которые мы приведём без доказательства со ссылкой на источники. Кроме того, мы сформулируем и докажем в виде лемм несколько важных вспомогательных утверждений.

§ 4. Очень важная лемма

О малой теореме Ферма и теореме Эйлера написано немало, см., например, подборку статей [1, 2, 7]. В этом разделе мы приведём формулировки теорем, а чуть позже применим теоремы при решении наших задач.

Теорема 1 (Малая теорема Ферма). *Если числа a и p взаимно просты и p простое, то $a^{p-1} \equiv 1 \pmod{p}$.*

При решении задач 1 и 2 мы столкнулись с необходимостью выяснить, на какую степень двойки делится число $3^m - 1$. Частичный ответ на этот вопрос мы можем почерпнуть из теоремы Эйлера.

Теорема 2 (Эйлер). *Если числа a и m взаимно просты, то $a^{\phi(m)} \equiv 1 \pmod{m}$, здесь $\phi(m)$ — функция Эйлера.*

Возьмём $m = 2^{n+1}$, тогда $\phi(m) = 2^n$. Поэтому $3^{2^n} \equiv 1 \pmod{2^{n+1}}$ и, более того, для любого натурального t выполняется $(3^t)^{2^n} \equiv 1 \pmod{2^{n+1}}$, поскольку числа 3^t и 2^{n+1} взаимно просты. Следовательно, число $3^{2^n t} - 1$ делится на 2^{n+1} и, конечно, делится на степени двойки с меньшими показателями.

И всё же в нашем случае ответ, полученный с помощью теоремы Эйлера, не является исчерпывающим. Фактически мы хотим узнать, на какую *максимальную* степень двойки делится число $3^m - 1$. Такой показатель степени 2 называется 2-адическим (диадическим) показателем для числа $3^m - 1$. Нам поможет следующая

Лемма 1. *Для $i \geqslant 1$ выполняется равенство $3^{2^i} = 1 + 2^{i+2} + 2^{i+3}q$, где $q \in \mathbb{Z}$.*

Доказательство.

$$\frac{3^{2^i}-1}{2^{i+2}} = \frac{(3-1)(3+1)(3^2+1)\dots(3^{2^{i-1}}+1)}{2^{i+2}} = \frac{3^2+1}{2}\cdot\frac{3^4+1}{2}\cdot\ldots\cdot\frac{3^{2^{i-1}}+1}{2} = r.$$

Каждый сомножитель представляет собой нечётное число, поскольку является половиной суммы квадрата нечётного числа и 1:

$$\frac{(2x+1)^2+1}{2} = 1 + 2(x^2+x)$$

— нечётное число. Следовательно, число r также нечётно: $r = 2q+1$, где $q \in \mathbb{Z}$. Тогда $\frac{3^{2^i}-1}{2^{i+2}} = 2q+1$, откуда $3^{2^i} = 1 + 2^{i+2} + 2^{i+3}q$, ч. т. д. □

Из леммы 1 вытекает очевидное

Следствие 1. *Для $i \geqslant 1$ выполняется*
а) $3^{2^i} \equiv 1 \pmod{2^{i+2}}$; б) $3^{2^i} \equiv 1 + 2^{i+2} \pmod{2^{i+3}}$.

Рассмотрим выражение $3^{2^n t} - 1$ для нечётного числа t. Мы можем разложить его на множители:

$$3^{2^n t} - 1 = (3^{2^n})^t - 1 = (3^{2^n}-1)\big((3^{2^n})^{t-1} + (3^{2^n})^{t-2} + \ldots + 3^{2^n} + 1\big).$$

Множитель $(3^{2^n})^{t-1} + (3^{2^n})^{t-2} + \ldots + 3^{2^n} + 1$ является нечётным числом как сумма нечётного числа нечётных слагаемых. Следовательно, максимальная степень двойки, на которую делится число $3^{2^n t} - 1$, совпадает с максимальной степенью двойки, на которую делится число $3^{2^n} - 1$. Таким образом, лемма улучшает результат, полученный из теоремы Эйлера, а именно: если t — нечётное число, то число $3^{2^n t} - 1$ делится на 2^{n+2} и не делится на 2^{n+3}. Другими словами, число $n+2$ является 2-адическим (диадическим) показателем числа $3^{2^n t} - 1$.

Фактически мы доказали следующее утверждение:

Лемма 2. *Если $3^m - 1$ делится на 2^a и не делится на 2^{a+1}, то $m = 2^{a-2}t$, где t — нечётное число.*

В такой формулировке этот результат мы будем часто применять в дальнейшем. Интересные примеры нахождения p-адических показателей для случая простого числа $p \geqslant 3$ читатели могут найти в статье [4].

Опираясь на лемму 2, мы можем получить другой способ решения задачи 2.

Решение задачи 2 (II способ). По условию $1 + 2^a + 2^b = 3^m$. Правая часть уравнения делится на три, следовательно, и левая часть делится

на три. Это возможно только в том случае, когда оба показателя степеней двоек — чётные числа. Значит, $a = 2a_1 \geqslant 2$ и $b = 2b_1 \geqslant 2$. Перепишем наше уравнение в виде

$$2^a(1 + 2^{b-a}) = 3^m - 1.$$

Поскольку $1 + 2^{b-a}$ — нечётное число, получаем, что $3^m - 1$ делится на 2^a и не делится на 2^{a+1}. Тогда из леммы 2 следует, что $m = 2^{a-2}t$, где t нечётно и $a = 2a_1 > 3$. Предположим, что $a > 5$. Тогда $3^{2^{a-3}t} \geqslant 3^{2^{a-3}} > 2^a$. Имеем

$$1 + 2^a = 3^m - 2^b = (3^{2^{a-3}t})^2 - (2^{b_1})^2 = (3^{2^{a-3}t} - 2^{b_1})(3^{2^{a-3}t} + 2^{b_1}).$$

Значит, $3^{2^{a-3}t} - 2^{b_1} \geqslant 1$. Далее,

$$1 + 2^a = (3^{2^{a-3}t} - 2^{b_1})(3^{2^{a-3}t} + 2^{b_1}) \geqslant 1 \cdot (3^{2^{a-3}t} + 2^{b_1}) > 2^a + 2^{b_1}.$$

Получаем, что $1 + 2^a > 2^a + 2^{b_1}$, что невозможно. В силу чётности a, остаётся только $a = 4$. Имеем

$$17 = 1 + 2^4 = 3^m - 2^b = (3^{2t} - 2^{b_1})(3^{2t} + 2^{b_1}),$$

откуда $1 = 3^{2t} - 2^{b_1}$ и $17 = 3^{2t} + 2^{b_1}$. Получаем $b = 6$, $m = 4$.

В итоге имеем единственное решение $(a; b; m) \in \{(4; 6; 4)\}$.

Упражнение 4. Найти все решения уравнений в целых неотрицательных числах a, b, m, где $1 \leqslant a < b$:

а) $1 + 2^a + 2^b = 5^m$; б) $1 + 2^a + 2^b = 7^m$; в)* $1 + 2^a + 2^b = 2015^m$.

После удачного рассмотрения нескольких частных случаев хочется попробовать решить задачу в целом.

§ 5. Сумма цифр

Из признаков делимости на 3 и на 9, изучаемых в школе, несложно вывести, что сумма цифр в десятичной записи числа 3^m равна 3 для $n = 1$ и будет делиться на 9 для $n \geqslant 2$.

А что мы можем сказать о сумме цифр числа 3^n в двоичной записи?

Перепишем уравнения (1), (3), (5) немного по-другому, поменяв местами правую и левую части:

$$3^m = 2^n + 2^0;$$
$$3^m = 2^b + 2^a + 2^0;$$
$$3^m = 2^{a_k} + \ldots + 2^{a_2} + 2^{a_1} + 2^0.$$

Мы записали степень тройки в двоичной системе. Суммы цифр в правой части записанных уравнений соответственно равны 2, 3 и $k+1$.

Поскольку мы решили задачу 1, то тем самым ответили на вопрос: «Какие степени числа 3 имеют сумму цифр в двоичной записи, равную 2?» Это будут только две степени тройки: 3 и 9. Соответственно решения задач 2 и 3 отвечают на вопрос: «Какие степени числа 3 имеют сумму цифр в двоичной записи равную 3 или 4?»

Вопрос, поставленный в исследовательской задаче 4, теперь можно переформулировать по-другому: *«Конечно ли количество степеней троек, у которых сумма цифр в двоичной записи равна некоторому постоянному числу k?»*

Сразу же возникает и другой вопрос: если конечно, то можно ли как-то оценить количество таких степеней троек, т. е. понять, сколько различных решений имеет диофантово уравнение (5)?

Если мы обозначим через $\nu(n)$ сумму цифр в двоичной записи числа n, то наш вопрос сводится к решению уравнения

$$\nu(3^m) = k. \tag{6}$$

Конечно, такая упрощённая запись не означает, что новое уравнение легче решается. Но порой, переформулировав задачу и посмотрев на неё с другой стороны, удаётся увидеть некоторые скрытые закономерности.

Такая переформулировка позволила нам подключить к исследованию задачи компьютер. Нужно было написать программу, которая записывает степени тройки в двоичной системе, считает сумму цифр и сводные данные заносит в таблицу. Школьник Влад Сергиенко написал такую программу, дополнив её графиками функции $y = \nu(3^x)$. На самом деле мы имеем дело с последовательностью $s_m = \nu(3^m)$, а программа по имеющейся последовательности значений построила кусочно-линейную функцию. Расчёты были произведены для всех целых $0 \leqslant m \leqslant 299$.

Приведём результаты этих вычислений, где q обозначает количество различных решений для различных значений суммы цифр s (см. таблицу на стр. 176).

Начало таблицы соответствует тому, что мы уже получили. Имеется только одна степень тройки с суммой цифр в двоичной записи, равной 1, — при $m = 0$. Только две степени тройки с суммой цифр в двоичной записи, равной 2, — при $m = 1$ и $m = 2$. Только одна степень тройки с суммой цифр в двоичной записи, равной 3, — при $m = 4$. По одной степени тройки с суммами цифр в двоичной записи, равными 4 и 5.

Расчёты также показали, что при $m \leqslant 299$ уравнение $\nu(3^m) = 6$ и, соответственно, уравнение $1 + 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} = 3^m$, где $1 \leqslant a_1 < a_2 <$

s	q	s	q	s	q	s	q	s	q	s	q	s	q	s	q
1	1	25	3	49	1	75	2	106	2	130	2	158	1	187	1
2	2	26	2	51	1	77	3	109	1	131	1	159	1	189	2
3	1	27	1	52	1	79	2	110	1	132	1	160	2	190	1
4	1	28	1	55	2	81	3	111	2	133	1	162	3	191	2
5	1	29	1	56	3	82	2	112	1	136	1	165	1	192	1
6	3	30	1	57	1	83	2	114	1	138	2	166	1	194	2
8	1	32	2	58	3	84	1	115	2	140	1	170	4	195	2
9	1	33	1	61	1	85	2	116	1	143	1	171	7	196	4
10	1	34	4	63	1	87	3	118	1	145	1	172	1	197	1
11	2	35	1	64	2	89	1	119	4	146	2	173	1	198	5
13	1	36	1	65	1	90	2	120	2	147	2	174	1	199	1
14	3	38	2	66	1	91	2	121	1	148	1	175	2	200	1
15	1	39	4	67	3	92	1	122	1	149	1	178	1	201	2
16	1	40	1	68	2	93	2	123	2	150	1	179	3	203	1
17	2	41	2	69	1	94	1	124	3	151	2	180	3	204	2
18	1	42	3	70	2	95	1	125	2	153	1	181	2	206	3
19	1	45	2	71	2	97	5	126	1	154	2	182	1	208	1
20	1	46	1	72	2	98	2	127	2	155	1	183	3	209	2
22	1	47	1	73	1	100	1	128	1	156	2	185	1	210	3
24	1	48	1	74	1	105	1	129	4	157	1	186	1	212	2

$< a_3 < a_4 < a_5$, имеет три решения. Уравнение $\nu(3^m) = 7$ и, соответственно, уравнение $1 + 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} = 3^m$, где $1 \leqslant a_1 < a_2 < a_3 < a_4 < < a_5 < a_6$, не имеет ни одного решения.

Последний факт явился неожиданностью, поскольку на первый взгляд казалось, что с увеличением числа слагаемых и, соответственно, количества независимых переменных решения всегда должны быть. Из таблицы также видно, что этот факт не единичный и, например, уравнения $\nu(3^m) = 12$ или $\nu(3^m) = 21$ также не имеют решений при $m \leqslant 299$. Взглянув на график функции $y = \nu(3^x)$ и на аналогичные графики по другим ос-

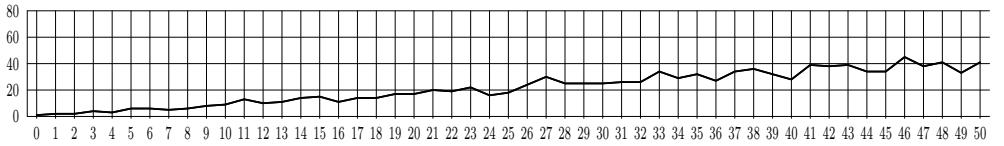


Рис. 1. Сумма цифр числа 3^n *по основанию* 2

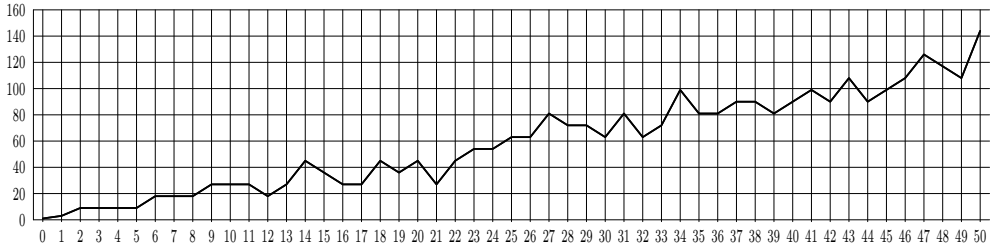


Рис. 2. Сумма цифр числа 3^n *по основанию* 10

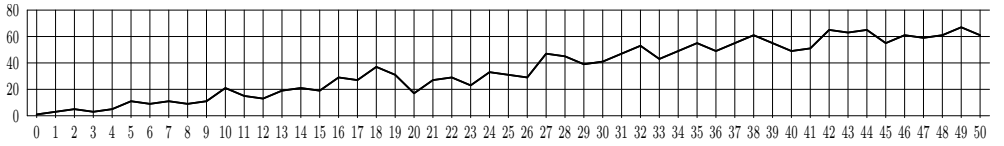


Рис. 3. Сумма цифр числа 3^n *по основанию* 5

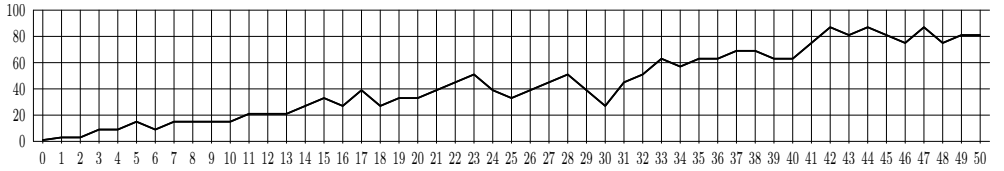


Рис. 4. Сумма цифр числа 3^n *по основанию* 7

нованиям систем счисления на расчётном интервале (рис. 1–4), мы видим функции, в среднем возрастающие с ростом рассматриваемого интервала.

Получив такие данные, мы вполне можем переформулировать наши вопросы в виде эквивалентных утверждений в следующей теореме (одно из них является ответом на вопрос задачи 4).

Теорема 3. (i) *Уравнение* $1 + 2^{a_1} + 2^{a_2} + 2^{a_3} + \ldots + 2^{a_k} = 3^m$, *где* $1 \leqslant a_1 < a_2 < a_3 < \ldots < a_k$, *при фиксированном натуральном* k *имеет конечное число решений.*

(ii) *Последовательность* $s_m = \nu(3^m)$ *возрастает неограниченно вместе с* m, *т. е.*

$$\varliminf_{m \to \infty} s_m = \infty.$$

Элементарное доказательство приведённой теоремы авторам неизвестно. Известно, что теорема может быть доказана разными, но, как мы отметили, не элементарными способами. Ввиду их сложности оставим решение этой теоремы за рамками статьи.

Следует добавить, что приведённая теорема является следствием серьёзной теоремы Шмидта о подпространстве (см. [15]). При таком подходе можно получить неэффективную оценку на рост суммы цифр. Эффективную оценку с помощью метода Бейкера получил Стюарт. Заинтересованный читатель может ознакомиться с указанными результатами, изучив источники [14, 16, 17, 20].

§ 6. О теореме Шинцеля

Если рассмотреть десятичную систему счисления, то аналог теоремы 3 удаётся доказать с помощью непростых, но вполне элементарных рассуждений.

Задача 5 [9, задача 207]. Доказать, что сумма цифр числа 2^n (записанного в десятичной системе счисления) неограниченно возрастает вместе с n.

На самом деле имеет место более сильное утверждение.

Теорема 4 (А. Шинцель). *Если g — чётное натуральное число, не делящееся на* 10*, то сумма цифр числа g^n (записанного в десятичной системе счисления) возрастает неограниченно вместе с n.*

Мы приведём доказательство теоремы А. Шинцеля по книге [9].

Доказательство. Определим бесконечную последовательность целых чисел a_i, $i = 0, 1, 2, \ldots$, следующим образом: $a_0 = 0$ и для $k = 0, 1, 2, \ldots$ пусть a_{k+1} означает наименьшее натуральное число, такое что $2^{a_{k+1}} > 10^{a_k}$ (таким образом, $a_1 = 1$, $a_2 = 4$, $a_3 = 14$ и т. д.). Имеем $a_0 < a_1 < a_2 < \ldots$

Докажем, что если при натуральном k выполнено $n \geqslant a_k$, то сумма цифр числа g^n будет не меньше чем k. Пусть c_j — цифра десятичного разложения числа g^n, стоящая при 10^j. Так как g есть чётное число, получаем, что $2^n \mid g^n$, и так как $n \geqslant a_k$, получаем $2^{a_i} \mid g^n$ для $i = 0, 1, 2, \ldots, k$. Поэтому, учитывая, что $2^{a_i} \mid 10^{a_i}$, имеем:

$$2^{a_i} \mid (c_{a_i-1}10^{a_i-1} + \ldots + c_0).$$

Если бы для $a_{i-1} \leqslant j < a_i$ все цифры c_j были равны нулю, то мы имели бы:

$$2^{a_i} \mid (c_{a_{i-1}-1}10^{a_{i-1}-1} + \ldots + c_0)$$

и, так как $c_0 \neq 0$, также

$$2^{a_i} \leqslant c_{a_i-1}10^{a_{i-1}-1} + \ldots + c_0 < 10^{a_{i-1}}.$$

Откуда $2^{a_i} \leqslant 10^{a_{i-1}}$, что противоречит определению числа a_i. Следовательно, по крайней мере одна из цифр c_j, где $a_{i-1} \leqslant j < a_i$, отлична от нуля.

Но последнее заключение справедливо для $i = 0, 1, 2, \ldots, k$. Следовательно, число g^n имеет по крайней мере k цифр отличных от нуля. Поэтому для достаточно больших n (для $n \geqslant a_k$) сумма цифр числа g^n не меньше произвольно заданного натурального числа k и, значит, сумма цифр числа g^n возрастает неограниченно вместе с n. Что и требовалось доказать. □

Упражнение 5 (А. Шинцель). Если g — нечётное натуральное число, делящееся на 5, то сумма цифр числа g^n (записанного в десятичной системе счисления) возрастает неограниченно вместе с n.

В частности, из доказанной теоремы следует для $g = 2$, что сумма цифр числа 2^n возрастает неограниченно вместе с n. Однако она не возрастает монотонно, например сумма цифр 2^3 есть 8, а сумма цифр 2^4 есть 7 или сумма цифр 2^{16} есть 25, а сумма цифр 2^{17} есть 14.

Для сравнения на рис. 5–8 приведены графики суммы цифр числа 2^n в системах счисления по основаниям 10, 3, 5, 7. Неограниченность роста

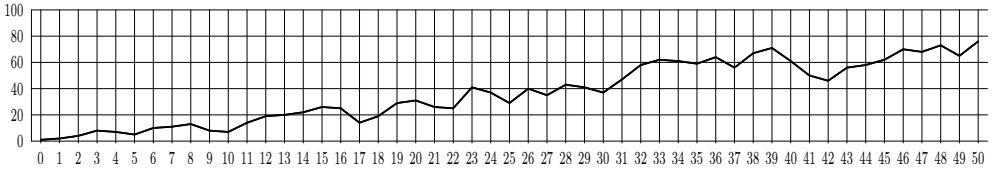


Рис. 5. Сумма цифр числа 2^n *по основанию* 10

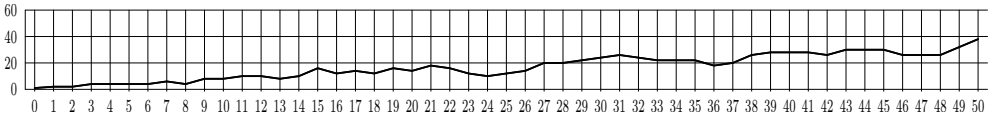


Рис. 6. Сумма цифр числа 2^n *по основанию* 3

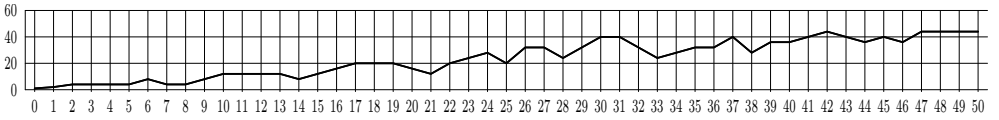


Рис. 7. Сумма цифр числа 2^n *по основанию* 5

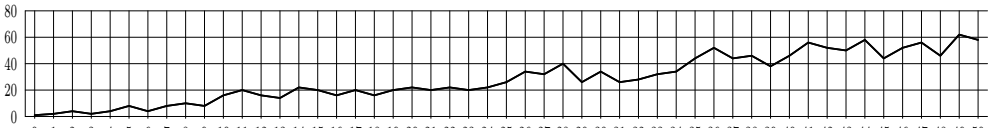


Рис. 8. Сумма цифр числа 2^n *по основанию* 7

суммы цифр числа 2^n по основанию 10 следует из теоремы Шинцеля. Для оснований 3, 5, 7 она следует из уже упоминавшейся теоремы Шмидта о подпространстве. Элементарное доказательство этого факта авторам неизвестно.

К сожалению, мы не смогли адаптировать имеющееся доказательство теоремы Шинцеля для случая теоремы 3. Дело в том, что в этом доказательстве существенную роль играет тот факт, что основание степени и основание системы счисления имеют общий делитель. В исследуемом нами случае числа 2 и 3 взаимно просты.

§ 7. В поисках формулы для суммы цифр

На каком-то этапе нам показалось, что утверждение (ii) теоремы 3 поддаётся элементарному доказательству и с ним можно «поработать». А что если найти явную формулу для вычисления суммы цифр числа?

Рассмотрим двоичную запись числа n. Пусть

$$n = b_q 2^q + b_{q-1}2^{q-1} + \ldots + b_1 2 + b_0, \quad \text{где } b_q = 1,\ b_i \in \{0, 1\}.$$

Как и прежде, $\nu(n)$ обозначает сумму цифр двоичной записи числа n, т. е. количество тех чисел b_i, которые равны 1. Понятно, что $q = [\log_2 n]$.

Через $\tau(n)$ обозначим максимальное количество единиц в конце двоичной записи числа n, т. е. $b_i = 1$, если $0 \leqslant i \leqslant \tau(n) - 1$, и $b_{\tau(n)} = 0$.

Задача 6. Докажите, что для всех натуральных n выполнены соотношения:

а) $\nu(n+1) = \nu(n) + 1 - \tau(n)$;

б) $\nu(n) = n - \sum\limits_{k=1}^{q} k\left[\dfrac{n+2^k}{2^{k+1}}\right]$, где $q = [\log_2 n]$.

Решение. а) Если n чётное, т. е. $b_0 = 0$ и поэтому $\tau(n) = 0$, имеем очевидное равенство $\nu(n+1) = \nu(n) + 1$.

Если n нечётное, то, поскольку у n на конце $\tau(n)$ единиц и $b_{\tau(n)} = 0$, то у $n+1$ на конце будет $\tau(n)$ нулей, при этом на $(\tau(n)+1)$-м месте появится 1, остальные цифры не изменятся. Следовательно, $\nu(n+1) = \nu(n) + 1 - \tau(n)$. Что и требовалось доказать.

б) Используем доказанное равенство. Имеем:

$$\begin{aligned} \nu(n) &= \nu(n-1) + 1 - \tau(n-1), \\ \nu(n-1) &= \nu(n-2) + 1 - \tau(n-2), \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{aligned}$$

$$\nu(2) = \nu(1) + 1 - \tau(1),$$
$$\nu(1) = 1.$$

Сложив все равенства, после упрощения получим

$$\nu(n) = n - \sum_{i=1}^{n-1} \tau(i).$$

Пусть t_k — количество чисел, меньших n и имеющих на конце k единиц. Нетрудно видеть, что числа вида $2^k - 1 + 2^{k+1}t$, где $t = 0, 1, 2, \ldots$, и только они, имеют на конце k единиц. Решив неравенство

$$2^k - 1 + 2^{k+1}t \leqslant n - 1,$$

получаем

$$t_k = 1 + \left[\frac{n - 2^k}{2^{k+1}}\right] = \left[\frac{n + 2^k}{2^{k+1}}\right].$$

Имеем

$$\nu(n) = n - \sum_{i=1}^{n-1} \tau(i) = n - \sum_{k=0}^{q} k\left[\frac{n + 2^k}{2^{k+1}}\right].$$

Что и требовалось доказать.

Теперь мы можем записать

$$\nu(3^m) = 3^m - \sum_{k=1}^{q} k\left[\frac{3^m + 2^k}{2^{k+1}}\right], \quad \text{где } q = [m \log_2 3].$$

Итак, формулу мы нашли!

Эта формула позволяет упростить процедуру программирования, позволяя избежать процедуры разложения степени тройки в двоичную запись. Однако такая формула не помогла ответить на ряд, казалось бы, простых вопросов. Например: «Какие степени числа 3 имеют сумму цифр в двоичной записи, равную 5?» Чтобы ответить на этот вопрос, приходится поступать как раньше, а именно решать в натуральных числах соответствующее экспоненциальное диофантово уравнение.

Искушённый читатель может встретить эту формулу в другом (возможно более простом) виде, восходящем к Лежандру (см., например, [18]):

$$s_b(n) = (b-1)\sum_{i=1}^{\infty}\left\{\frac{n}{b^i}\right\}, \quad \text{или} \quad s_b(n) = n - (b-1)\sum_{i=1}^{\infty}\left\lfloor\frac{n}{b^i}\right\rfloor,$$

где $s_b(n)$ обозначает сумму цифр b-ичной записи числа n.

На этом этапе мы вынуждены опять вернуться к частному случаю пока не разобранной задачи 3.

§ 8. Ещё несколько примеров

При попытках решить задачу 3 порой мы оказывались в ситуации уравнения (2). Поэтому для дальнейшего движения вернёмся к уравнению (2) и рассмотрим несколько частных случаев.

Задача 7. Решить уравнение $19 + 2^n = 3^m$ в натуральных числах.

Решение. Докажем, что у этого уравнения нет решений при $n \geqslant 5$. Действительно, если $n \geqslant 5$, то $m > 3$. Перепишем наше уравнение в виде

$$2^4(1 + 2^{n-4}) = 3(3^{m-1} - 1).$$

Поскольку $1 + 2^{n-4}$ — нечётное число, из леммы 2 получаем, что $m - 1 = 2^2(2t + 1) = 4(2t + 1)$. Кроме того, поскольку правая часть делится на 3, то и левая часть уравнения делится на 3, что возможно только при нечётном n. Имеем $19 + 2^n = 3 \cdot 81^{2t+1}$. Очевидно, что число в правой части равенства оканчивается цифрой 3. Однако поскольку n нечётно, то число в левой части может заканчиваться только цифрами 1 и 7. Получаем противоречие. Итак $n \leqslant 4$. Непосредственным перебором убеждаемся, что подходит только $n = 3$, тогда $m = 3$.

Ответ: $(n; m) \in \{(3; 3)\}$.

Задача 8. Решить уравнение $11 + 2^n = 3^m$ в натуральных числах.

Попробуем поступить так же, как и при решении предыдущей задачи.

Решение. Докажем, что у этого уравнения нет решений при $n \geqslant 5$. Действительно, если $n \geqslant 5$, то $m > 3$. Перепишем наше уравнение в виде

$$2^4(1 + 2^{n-4}) = 27(3^{m-3} - 1).$$

Поскольку $1 + 2^{n-4}$ — нечётное число, из леммы 1 получаем, что $m - 3 = 2^2(2t + 1) = 4(2t + 1)$. Имеем $11 + 2^n = 27 \cdot 81^{2t+1}$. Очевидно, что число в правой части равенства оканчивается цифрой 7. Число в левой части может заканчиваться на 7, только если 2^n оканчивается на 6. Следовательно, n делится на 4, т. е. $n = 4q$. Далее, $11 + 16^q = 27 \cdot 81^{2t+1}$. Однако $11 + 16^q \equiv 11 + (-1)^q \pmod{17}$. Значит, левая часть уравнения даёт остатки 10 или 12 при делении на 17. С другой стороны,

$$27 \cdot 81^{2t+1} \equiv 10 \cdot (-4)^{2t+1} \equiv -40 \cdot 16^t \equiv 11 \cdot (-1)^t \pmod{17}.$$

Значит, правая часть уравнения даёт остатки 6 или 11 при делении на 17. Получаем противоречие. Итак, $n \leqslant 4$. Непосредственным перебором убеждаемся, что подходит только $n = 4$, тогда $m = 3$.

Ответ: $(n; m) \in \{(4; 3)\}$.

Упражнение 6. Докажите, что следующие уравнения не имеют решений в натуральных числах m и n:

$$\text{а) } 35+2^n=3^m; \quad \text{б) } 37+2^n=3^m; \quad \text{в) } 41+2^n=3^m.$$

Упражнение 7. Решить уравнения в натуральных числах m и n:

$$\text{а) } 3+2^n=5^m; \quad \text{б) } 9+2^n=5^m; \quad \text{в) } 17+2^n=5^m; \quad \text{г) } 3+2^n=7^m;$$
$$\text{д) } 17+2^n=7^m; \quad \text{е) } 49+2^n=3^m; \quad \text{ж) } 65+2^n=3^m.$$

Как видим, при решении уравнений нам приходилось использовать несколько модулей. При этом если в задаче 7 нам хватило сравнений по модулю 10, то в задаче 8 нам дополнительно пришлось привлекать сравнения по модулю 17. Фактически мы решали системы сравнений.

§ 9. Системы сравнений

Следующая лемма выглядит несколько искусственно, и всё же она оказывается мощным орудием для решения задачи 3 и других подобных задач.

Лемма 3. *Система сравнений*

$$\begin{cases} 2^a+2^b+2^c\equiv 5 \pmod{15}, \\ 2^a+2^b+2^c\equiv 0 \pmod{17} \end{cases}$$

не имеет решений в целых неотрицательных числах a, b, c.

Потратив небольшие усилия, читатели смогут найти несколько различных доказательств этой леммы. Известные нам доказательства отличаются вариантами перебора остатков. Перебор вариантов можно сделать вручную или поручить компьютеру. Мы предложим вариант доказательства, в котором опустим рутинный перебор.

Доказательство. Предположим, что решение существует. Пусть $a = 4a_1 + r_1$, $b = 4b_1 + r_2$, $c = 4c_1 + r_3$, где $r_1, r_2, r_3 = 0, 1, 2, 3$. Имеем

$$2^a+2^b+2^c = 2^{r_1}16^{a_1}+2^{r_2}16^{b_1}+2^{r_3}16^{c_1} \equiv 2^{r_1}+2^{r_2}+2^{r_3} \pmod{15},$$

кроме того

$$2^a+2^b+2^c = 2^{r_1}16^{a_1}+2^{r_2}16^{b_1}+2^{r_3}16^{c_1} \equiv$$
$$\equiv 2^{r_1}(-1)^{a_1}+2^{r_2}(-1)^{b_1}+2^{r_3}(-1)^{c_1} \pmod{17}.$$

Если существует какое-нибудь решение, то очевидно, что существует решение, для которого $r_1 \leqslant r_2 \leqslant r_3$.

Это замечание сокращает перебор возможных вариантов. Перебрав их, получаем, что сравнение $2^{r_1} + 2^{r_2} + 2^{r_3} \equiv 5 \pmod{15}$ имеет решение в следующих двух случаях: 1) $0 = r_1 < r_2 = r_3 = 1$; 2) $2 = r_1 < r_2 = r_3 = 3$.

Разберём оба случая.

1) $0 = r_1 < r_2 = r_3 = 1$. Имеем

$$2^a + 2^b + 2^c \equiv (-1)^{a_1} + 2(-1)^{b_1} + 2(-1)^{c_1} \pmod{17}.$$

Поскольку $(-1)^{a_1} + 2(-1)^{b_1} + 2(-1)^{c_1} \equiv \pm 1, \pm 3, \pm 5 \pmod{17}$, в этом случае сравнение $2^a + 2^b + 2^c \equiv 0 \pmod{17}$ не выполняется.

2) $2 = r_1 < r_2 = r_3 = 3$. Имеем

$$2^a + 2^b + 2^c \equiv 4(-1)^{a_1} + 8(-1)^{b_1} + 8(-1)^{c_1} \equiv \\ \equiv 4((-1)^{a_1} + 2(-1)^{b_1} + 2(-1)^{c_1}) \pmod{17}.$$

Поскольку $4((-1)^{a_1} + 2(-1)^{b_1} + 2(-1)^{c_1}) \equiv \pm 3, \pm 4, \pm 5 \pmod{17}$, то и в этом случае сравнение $2^a + 2^b + 2^c \equiv 0 \pmod{17}$ не выполняется. Пришли к противоречию. Итак, система сравнений неразрешима. Что и требовалось доказать. □

Следствие 2. *Система сравнений*

$$\begin{cases} 2^a + 2^b \equiv 5 \pmod{15}, \\ 2^a + 2^b \equiv 0 \pmod{17} \end{cases}$$

не имеет решений в целых неотрицательных числах a, b.

Доказательство. Поскольку $a = b = 0$ не является решением системы, не теряя общности, можно считать $a \geqslant 1$. Тогда имеем эквивалентную систему

$$\begin{cases} 2^{a-1} + 2^{a-1} + 2^b \equiv 5 \pmod{15}, \\ 2^{a-1} + 2^{a-1} + 2^b \equiv 0 \pmod{17}. \end{cases}$$

Согласно лемме 3 последняя система сравнений не имеет решений. Что и требовалось доказать. □

Упражнение 8. Система сравнений

$$\begin{cases} 2^a + 2^b + 2^c \equiv 5 \pmod{15}, \\ 2^a + 2^b + 2^c \equiv 15 \pmod{17} \end{cases}$$

не имеет решений в целых неотрицательных числах a, b, c.

Упражнение 9. Системы сравнений

$$1)\ \begin{cases} 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}\equiv 170 \pmod{255},\\ 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}\equiv 0 \pmod{257}; \end{cases}$$

$$2)\ \begin{cases} 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}\equiv 170 \pmod{255},\\ 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}\equiv 0 \pmod{257} \end{cases}$$

не имеют решений в целых неотрицательных числах a_1, a_2, a_3, a_4, a_5.

Теперь мы готовы решить задачу 3.

§ 10. Уравнение $1+2^a+2^b+2^c=3^m$

Итак, решаем уравнение $1+2^a+2^b+2^c=3^m$, где $1\leqslant a<b<c$.

Решение задачи 3. Докажем, что уравнение $1+2^a+2^b+2^c=3^m$ не имеет решений в целых неотрицательных числах m, $a<b<c$ при $a\geqslant 6$. Действительно, если $a\geqslant 6$, то перепишем наше уравнение в виде

$$2^a(1+2^{b-a}+2^{c-a})=3^m-1.$$

Поскольку $1+2^{b-a}+2^{c-a}$ — нечётное число, из леммы 1 получаем, что $m=2^{a-2}q=16d$. Так как для любого натурального k выполняется $6^k\equiv 6 \pmod{15}$, получаем

$$3^m=3^{16d}=81^{4d}\equiv 6^{4d}\equiv 6 \pmod{15}.$$

Используя малую теорему Ферма, получим $3^{16}\equiv 1 \pmod{17}$, следовательно,

$$3^m=3^{16d}=(3^{16})^d\equiv 1^d\equiv 1 \pmod{17}.$$

Но тогда мы приходим к системе сравнений

$$\begin{cases} 2^a+2^b+2^c=3^m-1\equiv 5 \pmod{15},\\ 2^a+2^b+2^c=3^m-1\equiv 0 \pmod{17}. \end{cases}$$

Как следует из леммы 3, эта система не имеет решений. Итак, $a\leqslant 5$.

Разберём случай $a=2$. Поскольку $2=a<b<c$, получаем

$$5+2^b+2^c\equiv 5 \pmod{8}.$$

С другой стороны, 3^m при делении на 8 может дать остатки 1 и 3. Значит, в этом случае решений нет.

Случаи $a=1$, $a=3$, $a=4$, $a=5$ составляют содержание упражнений. Разобрав их, мы получим единственное решение $a=1$, $b=3$, $c=4$, $m=3$.

Ответ: $(a;b;c;m)\in\{(1;3;4;3)\}$.

Упражнение 10. Найти все решения уравнений в натуральных числах b, c, m, где $b < c$:

а) $3 + 2^b + 2^c = 3^m$; в) $17 + 2^b + 2^c = 3^m$;

б) $9 + 2^b + 2^c = 3^m$; г) $33 + 2^b + 2^c = 3^m$.

Упражнение 11. Доказать, что при $10 \leqslant a_1 < a_2 < a_3 < a_4 < a_5$ уравнение $1 + 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} = 3^m$ не имеет решений в натуральных числах.

§ 11. Простые числа Ферма

При решении нашей задачи потребовалось выбрать удобный модуль для сравнений. Простые модули, которые мы использовали, образуют начало одной из известных последовательностей — последовательности чисел Ферма:

$$F_n = 2^{2^n} + 1, n = 0, 1, 2, \ldots$$

Начальные члены этой последовательности следующие: 3, 5, 17, 257, 65 537, 4 294 967 297, ... В своё время Ферма считал, что все члены этой последовательности — простые числа. Однако в 1732 году Эйлер нашёл разложение

$$F_5 = 2^{2^5} + 1 = 4\,294\,967\,297 = 641 \cdot 6\,700\,417$$

и опроверг гипотезу Ферма. На сегодняшний момент известно лишь 5 простых чисел Ферма: 3, 5, 17, 257, 65 537; при этом существует гипотеза, что других простых чисел Ферма нет.

Однако вернёмся к нашей задаче. Давайте посмотрим на упражнение 9, в котором появляются модули $2^{2^3} - 1 = 255$ и $2^{2^3} + 1 = 257$. Используя те же рассуждения, что и при решении задачи 3, можно решить упражнение 10. На самом деле указанные модули работают и для большего числа слагаемых. С помощью компьютера авторы проверили, что верна

Теорема 5. *Система сравнений*

$$\begin{cases} 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} \equiv 170 \pmod{255}, \\ 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} \equiv 0 \pmod{257} \end{cases}$$

не имеет решений в целых неотрицательных числах a_i, $i = \overline{1, 7}$.

В то же время при выбранных модулях, увеличивая количество слагаемых, мы уже не приходим к противоречию.

Упражнение 12. Найдите хотя бы одно решение системы сравнений

$$\begin{cases} 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} + 2^{a_8} \equiv 170 \pmod{255}, \\ 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} + 2^{a_8} \equiv 0 \pmod{257}. \end{cases}$$

По всей видимости, чтобы получить результат, аналогичный теореме 5 или лемме 3 для восьми слагаемых (и, возможно, более восьми), мы должны взять модуль $2^{2^4} - 1 = 65\,535$ и модуль, соответствующий следующему простому числу Ферма $2^{2^4} + 1 = 65\,537$. Авторам не показалось возможным провести такую проверку на домашних компьютерах.

Гипотеза 1. *Система сравнений*

$$\begin{cases} 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} + 2^{a_8} \equiv 3^{65\,536} - 1 \pmod{65\,535}, \\ 2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} + 2^{a_6} + 2^{a_7} + 2^{a_8} \equiv 0 \pmod{65\,537} \end{cases}$$

не имеет решений в целых неотрицательных числах a_i, $i = \overline{1, 8}$.

Однако следующее число Ферма уже не является простым, и что будет происходить с системами сравнений в этом случае — авторам не ясно.

Может быть, читатели присоединятся к этому исследованию?

§ 12. Задачи для исследования

Задача 9. Решить уравнения в натуральных числах n, $m \neq 1$ и $q > 7$: а) $3 + 2^n = q^m$; б) $5 + 2^n = q^m$.

Задача 10. Решить уравнение $3^n + 100 = 7^m$ в натуральных числах.

Задача 11. Найти все решения уравнения $1 + 2^a + 2^b = y^2$ в целых неотрицательных числах a, b, y, где $1 \leqslant a < b$.

Задача 12. Найти все решения уравнения $1 + 2^a + 2^b = q^m$ в целых неотрицательных числах a, b, m, q, где $1 \leqslant a < b$, $m \geqslant 3$ и $q > 7$.

Задача 13 (Э. Б. Винберг, [5]). При каких натуральных n число $\dfrac{3^n - 1}{2}$ есть квадрат натурального числа?

В заключение отметим, что задача о росте суммы цифр успешно решалась не только для последовательности, состоящей из степеней натурального числа, но и для некоторых других последовательностей (см., например, [19]). Для искушённых читателей, вероятно, будет полезна книга [13].

§ 13. Ответы, указания, решения

Решения упражнений

Упражнение 1.

а) Ответ: $n = 2$, $m = 1$. б) Ответ: нет решений. в) Ответ: $q = 1 + 2^n$ и $m = 1$.

Решение. Решаем уравнение $1 + 2^n = q^m$. Очевидно, что q нечётно. Если m нечётно, то имеем $2^n = q^m - 1 = (q-1)(q^{m-1} + q^{m-2} + \ldots + q + 1)$. Множитель $q^{m-1} + q^{m-2} + \ldots + q + 1$ является суммой нечётного числа нечётных слагаемых и, следовательно, нечётен. Это возможно только при $m = 1$, тогда $q = 1 + 2^n$.

Если m чётно, то $m = 2t$. Тогда $2^n = (q^t - 1)(q^t + 1)$. Последнее равенство возможно, только если каждый из сомножителей в правой части является степенью двойки. Имеем

$$\begin{cases} q^t - 1 = 2^b, \\ q^t + 1 = 2^c \end{cases} \quad \Rightarrow \quad 2 = 2^c - 2^b \quad \Rightarrow \quad c = 2,\ b = 1,\ q = 3.$$

г) Ответ: нет решений. д) Ответ: $n = 3$, $m = 1$. е) Ответ: $q = 2^n - 1$ и $m = 1$.

Решение. Решаем уравнение $-1 + 2^n = q^m$. Очевидно, что q нечётно. Если m нечётно, то имеем

$$2^n = q^m + 1 = (q+1)(q^{m-1} - q^{m-2} + q^{m-3} - \ldots - q + 1).$$

Множитель $q^{m-1} - q^{m-2} + q^{m-3} - \ldots - q + 1$ является суммой нечётного числа нечётных слагаемых и, следовательно, нечётен. Это возможно только при $m = 1$, тогда $q = 2^n - 1$. Если m чётно, то $q^m \equiv 1 \pmod 4$. Если $n \geqslant 2$, то $-1 + 2^n \equiv 3 \pmod 4$. Противоречие. При $n = 1$ остаётся только $q = 1$.

Упражнение 2.

Ответ: $(x; y; z) \in \{(2; 2; 2)\}$.

Решение. Поскольку $3^x + 4^y \equiv 1 \pmod 3$ и $5^z \equiv (-1)^z \pmod 3$, то z чётно. Положим $z = 2t$. Перепишем наше уравнение в виде

$$3^x = 5^z - 4^y = (5^t - 2^y)(5^t + 2^y).$$

Тогда

$$\begin{cases} 5^t - 2^y = 3^b, \\ 5^t + 2^y = 3^c \end{cases} \quad \Rightarrow \quad 2^{y+1} = 3^c - 3^b = 3^b(3^{c-b} - 1) \quad \Rightarrow \quad b = 0.$$

С учётом результата задачи 1а) $c = 2$, $y = 2$, $t = 1$, $z = 2$, $x = 2$.

Упражнение 3.

Ответ: $(x; y) \in \{(0; -2), (0; 2), (4; -23), (4; 23)\}$.

Решение. Если $x < 0$, имеем $1 < 1 + 2^x + 2^{2x+1} < 1 + 1 + 2 = 4$, поэтому решений нет.

При $x = 0, 1, 2$ точный квадрат получается только при $x = 0$, $y = \pm 2$.

Пусть $x \geqslant 3$. Будем искать решения для $y \geqslant 0$ (пары $(x; y)$ и $(x; -y)$ входят в множество решений одновременно). Поскольку $(2^x)^2 = 2^{2x} < 1 + 2^x + 2^{2x+1} < 2^{2x} + 2^{2x} + 2^{2x+1} = (2^{x+1})^2$ и $y^2 = 1 + 2^x + 2^{2x+1}$, получаем $2^x < y < 2^{x+1}$.

Перепишем наше уравнение в виде $2^x(1 + 2^{x+1}) = (y - 1)(y + 1)$. Числа $y - 1$ и $y + 1$ — это последовательные чётные числа. Поэтому одно из них делится на 2 только в первой степени.

Рассмотрим два возможных случая.

В первом случае имеем

$$\begin{cases} y - 1 = 2^{x-1} \cdot t, \\ y + 1 = 2^{x-1} \cdot t + 2, \end{cases}$$

где t — нечётное число. Поскольку $2^x < y < 2^{x+1}$, а t нечётно, получаем $2^x < y = 2^{x-1} \cdot t + 1 < 2^{x+1}$, откуда $t = 3$. Тогда $1 + 2^x + 2^{2x+1} = (1 + 3 \cdot 2^{x-1})^2$.

Положим $z = 2^{x-1}$. Имеем $1 + 2z + 8z^2 = (1 + 3z)^2 \;\Leftrightarrow\; 0 = 4z + z^2$, т. е. в этом случае нет решений.

Аналогично во втором случае имеем

$$\begin{cases} y + 1 = 2^{x-1} \cdot t, \\ y - 1 = 2^{x-1} \cdot t - 2, \end{cases}$$

где t — нечётное число. Поскольку $2^x < y < 2^{x+1}$, а t — нечётное число, получим $2^x < y = 2^{x-1} \cdot t - 1 < 2^{x+1}$, откуда $t = 3$. Тогда $1 + 2^x + 2^{2x+1} = (3 \cdot 2^{x-1} - 1)^2$. Положим $z = 2^{x-1}$. Имеем $1 + 2z + 8z^2 = (3z - 1)^2 \;\Leftrightarrow\; 0 = z^2 - 8z$. Получаем $z = 8 = 2^{x-1}$. Откуда имеем ещё два решения $x = 4$ и $y = \pm 23$.

Упражнение 4.

а) Ответ: $(a; b; m) \in \{(3; 4; 2)\}$.

Решение. Если $b > a = 1$, то $1 + 2 + 2^b \equiv 3 \pmod 4$. С другой стороны, $5^m \equiv 1 \pmod 4$, что невозможно.

Если $a = 2$, то $5 + 2^b = 5^m$, что невозможно, поскольку 2^b не делится на 5. Итак, $a \geqslant 3$. Тогда $1 + 2^a + 2^b \equiv 1 \pmod 8$.

Если m — нечётное число, то $5^m \equiv 5 \pmod 8$ и в этом случае уравнение не имеет решений. Значит, m чётно: $m = 2k$. Имеем $2^a + 2^b = 25^k - 1$.

Пусть k нечётно. Имеем $2^a(1 + 2^{b-a}) = 25^k - 1 = 24(25^{k-1} + 25^{k-2} + \ldots + 1)$. Поскольку число $25^{k-1} + 25^{k-2} + \ldots + 1$ нечётно, как сумма нечётного числа нечётных слагаемых, получаем $a = 3$. Тогда $9 + 2^b = 25^k$. Так как $25^k \equiv 1 \pmod 3$, получаем, что и $2^b \equiv 1 \pmod 3$. Следовательно, b чётно: $b = 2t$. Тогда

$$9 = 25^k - 2^b = 25^k - 2^{2t} = (5^k - 2^t)(5^k + 2^t).$$

Это возможно, только если $k=1$ и $t=2$, тогда получаем решение $a=3$, $b=4$, $m=2$. Остаётся рассмотреть случай чётного $k \geqslant 2$. Положим $k=2k_1$. Далее, $25^k - 1 \equiv 0 \pmod 3$. С другой стороны, $2^a + 2^b \equiv (-1)^a + (-1)^b \pmod 3$. Следовательно, числа a, b имеют разную чётность. Кроме того, $1 + 2^a + 2^b \equiv 0 \pmod 5$. Это даёт нам только две возможности:

1) a делится на 4, т. е. $a = 4a_1$ и $b = 4b_1 + 3$;
2) $a = 4a_1 + 3$ и $b = 4b_1$.

В первом случае имеем

$$1 + 2^a + 2^b = 1 + 16^{a_1} + 8 \cdot 16^{b_1} \equiv 1 + (-1)^{a_1} + 8 \cdot (-1)^{b_1} \pmod{17}.$$

Аналогично во втором случае

$$1 + 2^a + 2^b = 1 + 8 \cdot 16^{a_1} + 16^{b_1} \equiv 1 + 8 \cdot (-1)^{a_1} + (-1)^{b_1} \pmod{17}.$$

Значит, и в первом, и во втором случае при делении на 17 левая часть уравнения $1 + 2^a + 2^b = 5^m$ может давать остатки $8, 9, 10, 11$. Однако, $25^k = 25^{2k_1} \equiv 8^{2k_1} \equiv (64)^{k_1} \equiv 13^{k_1} \pmod{17}$, и, следовательно, правая часть $1 + 2^a + 2^b = 5^m$ уравнения может давать остатки $1, 4, 13, 16$ при делении на 17, что невозможно. Следовательно, для чётных k решений нет.

Итак, мы нашли единственное решение $(a; b; m) \in \{(3; 4; 2)\}$.

б) Ответ: $(a; b; m) \in \{(1; 2; 1), (4; 5; 2)\}$.

Решение. Очевидно, что $a=1$, $b=2$, $m=1$ — решение уравнения.

Если $a=1$ и $b \geqslant 3$, то $1+2+2^b \equiv 3 \pmod 8$, с другой стороны, $7^m \equiv (-1)^m \pmod 8$, т. е. в этом случае нет решений.

Аналогично для $a=2$ и $b \geqslant 3$ имеем $1 + 4 + 2^b \equiv 5 \pmod 8$, что также невозможно.

Рассмотрим случай $a \geqslant 3$, тогда $1 + 2^a + 2^b \equiv 1 \pmod 8$. Как мы уже отметили, $7^m \equiv (-1)^m \pmod 8$, поэтому m должно быть чётным числом: $m = 2k$. Пусть $a = 4a_1 + r_1$, где $r_1 = 0, 1, 2, 3$, и $b = 4b_1 + r_2$, где $r_2 = 0, 1, 2, 3$. Имеем $1 + 2^a + 2^b = 1 + 2^{r_1}16^{a_1} + 2^{r_2}16^{b_1} \equiv 1 + 2^{r_1} + 2^{r_2} \pmod{15}$.

Занесём данные в таблицу остатков по модулю 15.

r_1	0	0	0	0	1	1	1	1	2	2	2	2	3	3	3	3
r_2	0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
$1 + 2^{r_1} + 2^{r_2}$	3	4	6	10	4	5	7	11	6	7	9	13	10	11	13	17

С другой стороны, $7^m = 7^{2k} = 49^k \equiv 4^k \pmod{15}$.

Если k — чётное число, т. е. $k=2k_1$, то $7^m \equiv 4^k = 4^{2k_1} = 16^{k_1} \equiv 1 \pmod{15}$, что невозможно, поскольку остаток 1 в таблице не встречается. Значит, k —

нечётно. В этом случае $7^m \equiv 4^k \equiv 4 \pmod{15}$. Далее, $1 + 2^a + 2^b = 7^m = 49^k$, $2^a(1 + 2^{b-a}) = 49^k - 1 = 48(49^{k-1} + 49^{k-2} + \ldots + 1)$.

Поскольку число $49^{k-1} + 49^{k-2} + \ldots + 1$ нечётно, как сумма нечётного числа нечётных слагаемых, получаем $a = 4$. Кроме того, как следует из таблицы остатков, $r_2 = 1$. Значит, $1 + 2^4 + 2^b = 49^k$. Перепишем это уравнение в виде $32(2^{b-5} - 1) = 49(49^{k-1} - 1)$. Для $k = 1$ получаем ещё одно решение $a = 4$, $b = 5$, $m = 2$.

Пусть теперь $k \geqslant 3$. Поскольку k нечётно, то $k - 1$ чётно. Положим $k - 1 = 2^d q$, где $q = 2q_1 + 1$ нечётно, а $d \geqslant 1$ — натуральное число. Тогда

$$49^{k-1} - 1 = 49^{2^d q} - 1 = (49^{2^d})^q - 1 = (49^{2^d} - 1)\big((49^{2^d})^{q-1} + (49^{2^d})^{q-2} + \ldots + 1\big).$$

Но $49^{2^d} - 1 = (49 - 1)(49 + 1)(49^2 + 1) \ldots (49^{2^{d-1}} + 1)$. Поэтому если $d \geqslant 2$, то правая часть уравнения делится на 64, а левая часть на 64 не делится. Значит, $d = 1$. Тогда $k = 2q + 1$ и $17 + 2^b = 49^{2q+1}$.

Рассмотрим остатки обеих частей уравнения по модулю 17. Имеем $17 + 2^b = 17 + 2^{r_2} 16^{b_1} = 17 + 2 \cdot 16^{b_1} \equiv 2 \cdot (-1)^{b_1} = \pm 2 \pmod{17}$. С другой стороны, $49^{2q+1} \equiv (-2)^{2q+1} \equiv (-2)^{4q_1+3} \equiv (-8) \cdot 16^{q_1} \equiv \pm 8 \pmod{17}$. Приходим к противоречию. Таким образом, других решений, кроме двух найденных, уравнение не имеет.

в) Ответ: нет решений.

Решение. Если $a = 1$, то $b > 4$, тогда $1 + 2 + 2^b \equiv 3 \pmod{16}$. С другой стороны $2015^m \equiv (-1)^m = \pm 1 \pmod{16}$. Противоречие.

Если $a = 2$, то $5 + 2^b = 2015^m$, что невозможно, так как 2^b не делится на 5.

Если $a = 3$ и $b > a$, то $9 + 2^b = 2015^m$. Но $9 + 2^b \equiv 9 \pmod{16}$ и $2015^m \equiv (-1)^m = \pm 1 \pmod{16}$. Противоречие.

Итак, $a \geqslant 4$. Тогда $1 + 2^a + 2^b \equiv 1 \pmod{16}$. Значит, m чётно, $m = 2t$. Поскольку 2015 делится на 5 и не делится на 3, правая часть уравнения даёт остаток 5 или 10 при делении на 15. Следовательно, левая часть уравнения также даёт остаток 5 или 10 при делении на 15. Как следует из таблицы остатков п. б), это возможно в трёх случаях:

1) $a = 4a_1$ и $b = 4b_1 + 3$;
2) $a = 4a_1 + 1$ и $b = 4b_1 + 1$;
3) $a = 4a_1 + 3$ и $b = 4b_1$.

Рассмотрим модуль 17. Имеем

$$2015^m \equiv 2015^{2t} \equiv 9^{2t} \equiv 81^t \equiv (-4)^t \equiv 1, 4, 13, 16 \pmod{17}.$$

В случае 1 имеем

$$1 + 2^a + 2^b = 1 + 16^{a_1} + 8 \cdot 16^{b_1} \equiv 1 + (-1)^{a_1} + 8(-1)^{b_1} \equiv 8, 9, 10, 11 \pmod{17},$$

что невозможно.

В случае 3 получаем такую же ситуацию.

В случае 2 получаем

$$1+2^a+2^b=1+2\cdot 16^{a_1}+2\cdot 16^{b_1}\equiv 1+2(-1)^{a_1}+2(-1)^{b_1}\equiv 1,5,14 \pmod{17}.$$

Возможен остаток 1, если a_1 и b_1 разной чётности, а $t=4t_1$. Пусть $a=8a_2+1$ и $b=8b_2+5$, случай $a=8a_2+5$ и $b=8b_2+1$ рассматривается аналогично. Имеем $1+2^{8a_2+1}+2^{8b_2+5}=2015^{8t_1}$. Далее

$$1+2^a+2^b=1+2\cdot 256^{a_2}+32\cdot 256^{b_2}\equiv 2 \pmod 3.$$

Но $2015^m\equiv 2015^{8t_1}\equiv 1 \pmod 3$. Поэтому решений нет.

Упражнение 5.

Доказательство. Определим бесконечную последовательность целых чисел a_i, $i=0,1,2,\ldots$, следующим образом: $a_0=0$ и для $k=0,1,2,\ldots$ пусть a_{k+1} означает такое наименьшее натуральное число, что $5^{a_{k+1}}>10^{a_k}$ (таким образом, $a_1=1$, $a_2=2$, $a_3=3$, $a_4=5$ и т. д.). Имеем $a_0<a_1<a_2<\ldots$

Докажем, что если при натуральном k выполнено $n\geqslant a_k$, то сумма цифр числа g^n будет $\geqslant k$. Пусть c_j — цифра десятичного разложения числа g^n, стоящая при 10^j. Так как g делится на 5, получаем, что $5^n\mid g^n$, и так как $n\geqslant a_k$, получаем, что $5^{a_i}\mid g^n$ для $i=0,1,2,\ldots,k$. Поэтому, учитывая, что $5^{a_i}\mid 10^{a_i}$, имеем: $5^{a_i}\mid (c_{a_i-1}\cdot 10^{a_i-1}+\ldots+c_0)$. Если бы для $a_{i-1}\leqslant j<a_i$ все цифры c_j были равны нулю, то мы имели бы: $5^{a_i}\mid (c_{a_{i-1}-1}\cdot 10^{a_{i-1}-1}+\ldots+c_0)$ и, так как $c_0\neq 0$, также $5^{a_i}\leqslant c_{a_{i-1}-1}\cdot 10^{a_{i-1}-1}+\ldots+c_0\leqslant 10^{a_{i-1}}$.

Откуда $5^{a_i}\leqslant 10^{a_{i-1}}$, что противоречит определению числа a_i. Следовательно, по крайней мере одна из цифр c_j, где $a_{i-1}\leqslant j<a$, отлична от нуля.

Но последнее заключение справедливо для $i=0,1,2,\ldots$ Следовательно, число g^n имеет по крайней мере k цифр отличных от нуля. Поэтому для достаточно больших n (для $n\geqslant a_k$) сумма цифр числа g^n не меньше произвольно заданного натурального числа k и, значит, сумма цифр числа g^n возрастает неограниченно вместе с n. Что и требовалось доказать. □

Упражнение 6.

а) При $n\leqslant 5$ решений нет. Если $n\geqslant 6$, то перепишем уравнение в виде

$$32(1+2^{n-5})=3(3^{m-1}-1).$$

Из леммы 2 получаем $m-1=2^{5-2}t=8t$, где t — нечётное число. Тогда $3^m=3^{8t+1}\equiv 3\cdot 81^{2t}\equiv 3\cdot 6\equiv 3 \pmod{15}$. С другой стороны, выражение $35+2^n$ при делении на 15 даёт остатки 6, 7, 9, 13. Противоречие.

б) При $n=1,2$ решений нет. Если $n\geqslant 3$, то $37+2^n\equiv 5 \pmod 8$, а $3^m\equiv 1,3 \pmod 8$, что невозможно.

в) При $n \leqslant 5$ решений нет. Если $n \geqslant 6$, то перепишем уравнение в виде

$$32(1 + 2^{n-5}) = 9(3^{m-2} - 1).$$

Из леммы 2 получаем $m - 2 = 2^{5-2}t = 8t$, где t — нечётное число. Тогда $3^m = 3^{8t+2} \equiv 9 \cdot 81^{2t} \equiv 9 \cdot 6 \equiv 9 \pmod{15}$. С другой стороны, выражение $41 + 2^n$ при делении на 15 даёт остатки $0, 4, 12, 13$. Противоречие.

Упражнение 7.

а) Ответ: $n = 1$, $m = 1$.

Если $n \geqslant 2$, то $3 + 2^n \equiv 3 \pmod 4$, а $5^m \equiv 1 \pmod 4$, что невозможно. Остаётся $n = 1$, $m = 1$.

б) Ответ: $n = 4$, $m = 2$.

При $n = 1, 2, 3$ решений нет. Рассмотрим случай $n \geqslant 4$.

Если m нечётно, то $8(1 + 2^{n-3}) = 5^m - 1 = 4(5^{m-1} + 5^{m-2} + \ldots + 1)$. Множитель $5^{m-1} + 5^{m-2} + \ldots + 1$ является нечётным числом, как сумма нечётного числа нечётных слагаемых. Поэтому правая часть уравнения делится на 4 и не делится на 8. Левая часть уравнения делится на 8. Противоречие.

Если m чётно, $m = 2t$, то $2^n = 5^m - 9 = (5^t - 3)(5^t + 3)$. Имеем для некоторых натуральных b, c:

$$\begin{cases} 5^t - 3 = 2^b, \\ 5^t + 3 = 2^c \end{cases} \quad \Rightarrow \quad 6 = 2^c - 2^b \quad \Rightarrow \quad c = 3,\ b = 1,\ t = 1,$$

откуда $m = 2$, $n = 4$. При решении можно также использовать результат упражнения 4а).

в) Ответ: $n = 3$, $m = 2$.

Указание. Перепишите уравнение в виде $1 + 2^4 + 2^n = 5^m$. Рассмотрите два случая $n \leqslant 4$ и $n > 4$. Используйте результат упражнения 4а).

г) Ответ: $n = 2$, $m = 1$. При $n = 1$ решений нет. При $n = 2$ получаем $m = 1$. Если $n \geqslant 3$, то $3 + 2^n \equiv 3 \pmod 8$, а $7^m \equiv \pm 1 \pmod 8$. Противоречие.

д) Ответ: $n = 5$, $m = 2$.

Указание. Перепишите уравнение в виде $1 + 2^4 + 2^n = 7^m$. Рассмотрите два случая $n \leqslant 4$ и $n \geqslant 4$. Используйте результат упражнения 4б).

е) Ответ: $n = 5$, $m = 4$. При $n = 1, 2$ решений нет. Пусть $n \geqslant 3$. Тогда $49 + 2^n \equiv 1 \pmod 8$, получаем $m = 2m_1$. Перепишем уравнение в виде $2^n = 3^m - 49 = (3^{m_1} - 7)(3^{m_1} + 7)$. Тогда

$$\begin{cases} 3^{m_1} - 7 = 2^x, \\ 3^{m_1} + 7 = 2^y \end{cases} \quad \Rightarrow \quad 14 = 2^y - 2^x \quad \Rightarrow \quad x = 1,\ y = 4 \quad \Rightarrow \quad m_1 = 2,\ m = 4.$$

Получаем $n = 5$.

ж) Ответ: $n=4$, $m=4$. При $n \leqslant 6$ получим решение перебором. Если $n \geqslant 7$, то перепишем уравнение в виде $64(1+2^{n-6}) = 3^m - 1$. Из леммы 2 получаем $m = 2^{6-2}t = 16t$, где t — нечётное число. Имеем $3^{16t} - 1 = 81^{4t} - 1 \equiv$ $\equiv 6 - 1 \equiv 5 \pmod{15}$ и по малой теореме Ферма $3^{16t} \equiv 1 \pmod{17}$. Имеем систему сравнений

$$\begin{cases} 2^6 + 2^n \equiv 5 \pmod{15}, \\ 2^6 + 2^n \equiv 0 \pmod{17}. \end{cases}$$

По следствию 2 эта система не имеет решений.

Упражнение 8.

Указание. Пусть $a = 4a_1 + r_1$, $b = 4b_1 + r_2$, $c = 4c_1 + r_3$, где $r_1, r_2, r_3 =$ $= 0, 1, 2, 3$. Имеем

$$2^a + 2^b + 2^c = 2^{r_1}16^{a_1} + 2^{r_2}16^{b_1} + 2^{r_3}16^{c_1} \equiv 2^{r_1} + 2^{r_2} + 2^{r_3} \pmod{15}$$

и

$$2^a + 2^b + 2^c \equiv 2^{r_1}(-1)^{a_1} + 2^{r_2}(-1)^{b_1} + 2^{r_3}(-1)^{c_1} \pmod{17}.$$

В доказательстве леммы 3 показано, что если $r_1 \leqslant r_2 \leqslant r_3$, то сравнение $2^a + 2^b + 2^c \equiv 5 \pmod{15}$ имеет решение в следующих двух случаях:

1) $0 = r_1 < r_2 = r_3 = 1$; 2) $2 = r_1 < r_2 = r_3 = 3$.

Тогда

$$2^a + 2^b + 2^c \equiv 2^{r_1}(-1)^{a_1} + 2^{r_2}(-1)^{b_1} + 2^{r_3}(-1)^{c_1} \in \{\pm 1, \pm 3, \pm 4, \pm 5\} \pmod{17}.$$

В этих случаях сравнение $2^a + 2^b + 2^c \equiv 15 \pmod{17}$ не выполняется.

Упражнение 9.

Отсутствие решений системы 1 следует из отсутствия решений системы 2. Действительно, в любом решении системы 1 какое-то из a_i было бы больше 0, но тогда можно заменить 2^{a_i} на $2^{a_i-1} + 2^{a_i-1}$, получив решение системы 2. Поэтому рассмотрим систему 2.

Пусть $a_i = 8b_i + r_i$, где $r_i = \overline{0,7}$. Имеем

$$2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} = \\ = 2^{r_1} \cdot 256^{b_1} + 2^{r_2} \cdot 256^{b_2} + 2^{r_3} \cdot 256^{b_3} + 2^{r_4} \cdot 256^{b_4} + 2^{r_5} \cdot 256^{b_5}.$$

Отсюда

$$2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} \equiv 2^{r_1} + 2^{r_2} + 2^{r_3} + 2^{r_4} + 2^{r_5} \pmod{255}$$

и

$$2^{a_1} + 2^{a_2} + 2^{a_3} + 2^{a_4} + 2^{a_5} \equiv 2^{r_1} \cdot (-1)^{b_1} + 2^{r_2} \cdot (-1)^{b_2} + \\ + 2^{r_3} \cdot (-1)^{b_3} + 2^{r_4} \cdot (-1)^{b_4} + 2^{r_5} \cdot (-1)^{b_5} \pmod{257}.$$

Если существует какое-нибудь решение, то очевидно, что существует решение, для которого $r_1 \leqslant r_2 \leqslant r_3 \leqslant r_4 \leqslant r_5$.

Проверьте, что все такие решения сравнения $2^{r_1} + 2^{r_2} + 2^{r_3} + 2^{r_4} + 2^{r_5} \equiv 170$ $\pmod{255}$ исчерпываются следующими случаями:

$r_1 = r_2 = 0$, $r_3 = 3$, $r_4 = 5$, $r_5 = 7$;
$r_1 = 1$, $r_2 = r_3 = 2$, $r_4 = 5$, $r_5 = 7$;
$r_1 = 1$, $r_2 = 3$, $r_3 = r_4 = 4$, $r_5 = 7$;
$r_1 = 1$, $r_2 = 3$, $r_3 = 5$, $r_4 = r_5 = 6$.

(Для сокращения перебора можно дополнительно использовать модули 3, 5 и 17.) Остаётся проверить, что эти решения не являются решениями сравнения

$$2^{r_1}\cdot(-1)^{b_1}+2^{r_2}\cdot(-1)^{b_2}+2^{r_3}\cdot(-1)^{b_3}+2^{r_4}\cdot(-1)^{b_4}+2^{r_5}\cdot(-1)^{b_5}\equiv 0 \pmod{257}.$$

Упражнение 10.

а) Ответ: $(b; c; m) \in \{(1; 2; 2), (3; 4; 3)\}$. Если $b = 1$, то мы получаем уравнение $5 + 2^c = 3^m$. Поскольку $3^m \equiv 1, 3 \pmod 8$, уравнение имеет решение только при $c = 2$.

Если $b = 2$, то $c \geqslant 3$ и мы получаем уравнение $7 + 2^c = 3^m$. Аналогично, поскольку $3^m \equiv 1, 3 \pmod 8$, а $7 + 2^c \equiv 7 \pmod 8$, в этом случае нет решений.

Если $b=3$ или $b=4$, то мы получим уравнения $11+2^c=3^m$ и $19+2^c=3^m$, которые рассмотрены в задачах 8 и 7.

Пусть теперь $5 \leqslant b < c$. Перепишем наше уравнение в виде

$$2^b(1 + 2^{c-b}) = 3(3^{m-1} - 1).$$

По лемме 2 получаем $m-1=2^{b-2}t=8d$, где t нечётно. Так как $3(3^{m-1}-1)=$ $=3(81^{2d}-1)\equiv 0 \pmod{15}$, правая часть уравнения делится на 15. В то же время левая часть уравнения на 15 не делится, поскольку $1+2^{c-b}\equiv 2, 3, 5, 9 \pmod{15}$. Следовательно, при $5 \leqslant b < c$ решений нет.

б) Ответ: $(b; c; m) \in \{(1; 4; 3), (3; 6; 4)\}$. Если $b = 1$, то мы получаем уравнение $11 + 2^c = 3^m$, которое рассмотрено в задаче 8.

Если $b=2$, то $c\geqslant 3$ и $13+2^c=3^m$. Так как $3^m\equiv 1,3 \pmod 8$, а $13+2^c\equiv 5 \pmod 8$, в этом случае нет решений.

Если $b = 3$, то $c \geqslant 4$ и $17 + 2^c = 3^m$. Поскольку правая часть уравнения делится на 3, то и левая должна делиться на 3, откуда следует, что $c=2c_1$ — чётное число. Поскольку $17+2^c\equiv 1 \pmod 8$, то $m=2m_1$. Перепишем наше уравнение в виде $17 = 3^m - 2^c = (3^{m_1} - 2^{c_1})(3^{m_1} + 2^{c_1})$. Тогда $17 = 3^{m_1} + 2^{c_1}$ и $1 = 3^{m_1} - 2^{c_1}$. Отсюда $m_1 = 2$, $c_1 = 3$ и $m = 4$, $c = 6$.

Пусть теперь $4 \leqslant b < c$. Перепишем наше уравнение в виде

$$2^b(1 + 2^{c-b}) = 9(3^{m-2} - 1).$$

По лемме 2 получаем $m-2=2^{b-2}t=4d$, где t нечётно. Так как $9(3^{m-2}-1)=$ $=9(81^d-1)\equiv 0 \pmod{15}$, правая часть уравнения делится на 15. В то же

время левая часть уравнения на 15 не делится, поскольку $1+2^{c-b} \equiv 2, 3, 5, 9 \pmod{15}$. Следовательно, при $4 \leqslant b < c$ решений нет.

в) Ответ: $(b; c; m) \in \{(1; 3; 3)\}$.

Если $b=1$, то мы получаем уравнение $19+2^c=3^m$, которое рассмотрено в задаче 7.

Если $b=2$, $b=4$ или $b=6$, то мы получаем уравнения $21+2^c=3^m$, $33+2^c=3^m$ и $81+2^c=3^m$. Эти уравнения не имеют решений, поскольку их правая часть делится на 3, а левая не делится.

Если $b=3$, то $c \geqslant 4$ и $25+2^c=3^m$. Так как $25+2^c \equiv 1 \pmod{8}$, то $m=2m_1$. Перепишем наше уравнение в виде $2^c=3^m-25=(3^{m_1}-5)(3^{m_1}+5)$. Тогда

$$\begin{cases} 3^{m_1}-5=2^x, \\ 3^{m_1}+5=2^y \end{cases} \quad \Rightarrow \quad 10=2^y-2^x.$$

Это уравнение не имеет решений в целых неотрицательных числах.

Если $b=5$, то $c \geqslant 6$ и мы получим уравнение $49+2^c=3^m$. Из упражнения 7е) следует, что оно имеет решение только при $c=5$. Противоречие.

Пусть теперь $7 \leqslant b < c$. Перепишем наше уравнение в виде

$$2^4(1+2^{b-4}+2^{c-4})=3^m-1.$$

Поскольку $1+2^{b-4}+2^{c-4}$ — нечётное число, по лемме 2 получаем $m=2^{4-2}t=4t$, где t нечётно. Положим $t=2s+1$.

Далее, $3^m-1=81^t-1 \equiv 5 \pmod{15}$, откуда $2^4+2^b+2^c \equiv 5 \pmod{15}$. Как следует из доказательства леммы 3, это возможно, только если $b \equiv c \equiv 1 \pmod{4}$, т. е. $b=4b_1+1$, $c=4c_1+1$. Имеем

$$16+2\cdot 2^{4b_1}+2\cdot 2^{4c_1}=81^{2s+1}-1=80\cdot(81^{2s}+81^{2s-1}+81+1),$$

тогда

$$1+2\cdot 16^{b_1-1}+2\cdot 2^{c_1-1}=5\cdot(81^{2s}+81^{2s-1}+81+1).$$

Заметим, что $1+2\cdot 16^{b_1-1}+2\cdot 2^{c_1-1} \equiv 1 \pmod{16}$. С другой стороны $5\cdot(81^{2s}+81^{2s-1}+81+1) \equiv 5\cdot(1^{2s}+1^{2s-1}+1+1)1 \equiv 5(2s+1) \pmod{16}$, поэтому $10s+5 \equiv 1 \pmod{16}$, т. е. $s \equiv 6 \pmod{8}$. Положим $s=8s_1+6$. Тогда $17+2\cdot 16^{b_1}+2\cdot 16^{c_1}=81^{16s_1+13}$. Поскольку

$$17+2\cdot 16^{b_1}+2\cdot 16^{c_1} \equiv 2\cdot(-1)^{b_1}+2\cdot(-1)^{c_1} \pmod{17}$$

и

$$81^{16s_1+13} \equiv 3^{4(16s_1+13)} \equiv 3^4\cdot 3^{16(4s_1+3)} \equiv 81 \equiv -4 \pmod{17},$$

получаем, что b_1 и c_1 — нечётные числа. Положим $b_1=2b_2+1$, $c_1=2c_2+1$. Тогда имеем $17+32\cdot 16^{2b_2}+32\cdot 16^{2c_2}=3^{64s_1+52}$ или $17+32\cdot 256^{b_2}+32\cdot 256^{c_2}=3^{64s_1+52}$.

Применим модуль 257.

Поскольку

$$17 + 32 \cdot 256^{b_2} + 32 \cdot 256^{c_2} \equiv 17 + 32 \cdot (-1)^{b_2} + 32 \cdot (-1)^{c_2} \pmod{257},$$

левая часть уравнения при делении на 257 даёт остатки $17, 81, 210$.

Пусть $s_1 = 2s_2 + r$, где $r = 0, 1$. Тогда получаем

$$3^{64s_1+52} = 3^{52} \cdot (3^{64})^{s_1} = 3^{52} \cdot (3^{64})^{2s_2+r} = \\ = 3^{52} \cdot (3^{64})^r \cdot (3^{128})^{s_2} \equiv 162 \cdot 241^r \cdot (-1)^{s_2} \pmod{257}.$$

Следовательно, правая часть уравнения при делении на 257 может давать остатки 22, 95, 162 и 235. Противоречие, т. е. при $7 \leqslant b < c$ нет решений.

г) Ответ: $(b; c; m) \in \{(4; 5; 4)\}$.

Если $b = 1$, $b = 2$ или $b = 3$, то мы получаем уравнения $35 + 2^c = 3^m$, $37 + 2^c = 3^m$, $41 + 2^c = 3^m$, которые не имеют решений (см. упражнение 6).

Если $b = 4$, то $49 + 2^c = 3^m$. Из упражнения 7е) следует, что оно имеет решение только при $c = 5$.

Если $b = 5$, то $c \geqslant 6$ и $65 + 2^c = 3^m$. Из упражнения 7ж) следует, что оно имеет решение только при $c = 4$. Противоречие.

Пусть теперь $6 \leqslant b < c$. Перепишем наше уравнение в виде

$$2^5(1 + 2^{b-5} + 2^{c-5}) = 3^m - 1.$$

По лемме 2 получаем $m = 2^{5-2}t = 8t$, где t — нечётное число, положим $t = 2s + 1$. Имеем $33 + 2^b + 2^c = 3^{16s+8}$, тогда $3^{16s+8} \equiv 81^{4s+2} \equiv 6 \pmod{15}$ и $3^{16s+8} \equiv 3^8 \cdot 3^{16s} \equiv 81^2 \equiv -1 \pmod{17}$. Получаем систему сравнений

$$\begin{cases} 2^5 + 2^b + 2^c \equiv 5 & \pmod{15}, \\ 2^5 + 2^b + 2^c \equiv -2 \equiv 15 & \pmod{17}. \end{cases}$$

Из упражнения 8 следует, что эта система не имеет решений. Следовательно, при $6 \leqslant b < c$ решений нет.

Упражнение 11.

Если $a_1 \geqslant 10$, то перепишем наше уравнение в виде

$$2^{a_1}(1 + 2^{a_2-a_1} + 2^{a_3-a_1} + 2^{a_4-a_1} + 2^{a_5-a_1}) = 3^m - 1.$$

Поскольку в скобках стоит нечётное число, из леммы 2 получаем, что $m = 2^{a_1-2}q = 256 \cdot d$.

Поскольку для любого натурального k выполняется $3^{16k} \equiv 171^k \equiv 171 \pmod{255}$, получаем, что $3^m = 3^{256d} \equiv 171 \pmod{255}$. Используя малую теорему Ферма, получим $3^m = 3^{256d} \equiv 1 \pmod{257}$. Но тогда мы приходим

к системе сравнений

$$\begin{cases} 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}=3^m-1\equiv 170 \pmod{255}, \\ 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}=3^m-1\equiv 0 \pmod{257}. \end{cases}$$

Как следует из упражнения 8, эта система не имеет решений, ч. т. д.

Упражнение 12.

В качестве одного из решений можно взять следующее: $a_1=0$, $a_2=8$, $a_3=2$, $a_4=10$, $a_5=4$, $a_6=12$, $a_7=6$, $a_8=14$.

Имеем

$$2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}+2^{a_6}+2^{a_7}+2^{a_8}=(2^0+2^8)(1+4+16+64)=257\cdot 85.$$

Следовательно, выполнено

$$\begin{cases} 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}+2^{a_6}+2^{a_7}+2^{a_8}\equiv 170 \pmod{255}, \\ 2^{a_1}+2^{a_2}+2^{a_3}+2^{a_4}+2^{a_5}+2^{a_6}+2^{a_7}+2^{a_8}\equiv 0 \pmod{257}. \end{cases}$$

Комментарии к задачам 9–13

9. По всей видимости, решений при $m>1$ и $q>7$ не существует, но доказательство этого факта авторам неизвестно.

10. Как мы обсуждали в § 2, уравнение имеет конечное число решений. Нетрудно найти пару $(n;m)=(5;3)$, являющуюся решением. Вероятно, других решений нет, но доказательство этого факта авторам неизвестно.

11. Задача является обобщением упражнения 3. Общее решение авторам неизвестно.

12. Задача обобщает идею задачи 3. Общее решение авторам неизвестно.

13. Ответ: $n\in\{1,2,5\}$. Это задача с историей. В своё время Э. Б. Винберг задал этот вопрос Ж.-П. Серру. Через несколько дней Серр прислал решение, основанное на p-адическом анализе. Решение задачи двумя способами, принадлежащее Н. Н. Осипову, опубликовано в [6]. Одно из решений, использующее теорию уравнений Пелля, доступно школьникам.

Благодарности

Авторы благодарят членов редколлегии «Математического просвещения» за ценные замечания.

Список литературы

[1] *Винберг Э. Б.* Малая теорема Ферма и её обобщения // Математическое просвещение. Сер. 3. Вып. 12. М.: МЦНМО, 2008. С. 43–53.

[2] *Гиндикин С. Г.* Малая теорема Ферма // Квант. 1972. № 10. С. 2–9.

[3] Задачник Кванта. 1978. № 6. М509 а); 1983. № 3. М792 а), б).

[4] *Кожевников П. А., Сендеров В. А.* Степени n и n-е степени // Квант. 2012. № 1. С. 9–12.

[5] Математическое просвещение. Сер. 3. Вып. 16. М.: МЦНМО, 2012. С. 231, задача 11.

[6] Математическое просвещение. Сер. 3. Вып. 17. М.: МЦНМО, 2013. С. 203–207.

[7] *Сендеров В. А., Спивак А. В.* Малая теорема Ферма // Квант. 2000. № 1. С. 9–16, 37; № 3. С. 11–17; № 4. С. 15–18.

[8] *Сендеров В. А., Френкин Б. Р.* Гипотеза Каталана // Квант. 2007. № 4. С. 8–10.

[9] *Серпинский В.* 250 задач по элементарной теории чисел. М.: Просвещение, 1968. Задачи 154, 155, 207.

[10] *Спивак А. В.* Арифметика-2 // Библиотечка Квант. Вып. 109. М.: Бюро Квантум, 2008. С. 90, упр. 64.

[11] *Фёдоров Р. М., Канель-Белов А. Я., Ковальджи А. К., Ященко И. В.* Московские математические олимпиады 1993–2005 г. М.: МЦНМО, 2006.

[12] *Ферма П.* Исследования по теории чисел и диофантову анализу. М.: Наука, 1992.

[13] *Shorey T. N., Tijdeman R.* Exponential diophantine equations. Cambridge: Cambridge University Press, 2008. (Cambridge Tracts in Math.; V. 87).

[14] *Stewart C. L.* On the representation of an integer in two different bases // J. Reine Angew. Math. 1980. V. 319. P. 63–72.

[15] http://en.wikipedia.org/wiki/Subspace_theorem.

[16] http://gdz.sub.uni-goettingen.de/dms/load/img/?PPN=GDZPPN002197707.

[17] http://mathoverflow.net/questions/143263/letting-sm-be-the-digit-sum-of-m-then-lim-n-to-inftys3n-infty.

[18] http://mathoverflow.net/questions/161960/looking-for-a-reference-to-a-classical-formula-for-the-sum-of-the-base-b-digit.

[19] http://mathoverflow.net/questions/179782/optimal-lower-bounds-for-the-sum-of-digits-in-base-b.

[20] http://mathoverflow.net/questions/38971/sums-of-digits-of-powers-of-integers.

В. М. Журавлёв, ОАО «Туполев», Москва

zhuravlevvm@mail.ru

П. И. Самовол, Университет им. Бен-Гуриона, проект Эйлера (Беэр-Шева, Израиль)

pet12@012.net.il

Разрезания прямоугольника на прямоугольники с заданными отношениями сторон

Ф. А. Шаров

§ 1. Основной результат

Рассмотрим такую задачу: *дан набор прямоугольников; какие фигуры можно разрезать на прямоугольники, подобные данным?* Поставленная задача в целом сложна, до сих пор не имеет решения и вряд ли может быть в разумном смысле решена в общем виде. Однако интересно было бы исследовать её частные случаи. В нашей статье мы выясним, какие прямоугольники можно разрезать на подобные n данным при условии, что отношения сторон данных прямоугольников — квадратичные иррациональности.

Сформулируем основную теорему нашей статьи. Этот результат — новый, причём доказательство использует только элементарные методы.

Теорема 1 (основная). *Пусть* $x_1 = a_1 + b_1\sqrt{p}$, $\ldots$, $x_n = a_n + b_n\sqrt{p}$ — *такие числа, что* $x_i > 0$, $a_i, b_i, p \in \mathbb{Q}$ $(1 \leqslant i \leqslant n)$ *и* $\sqrt{p} \notin \mathbb{Q}$. *Тогда*:

1) *если существуют такие числа* i *и* j, *что* $1 \leqslant i, j \leqslant n$ *и*

$$(a_i - b_i\sqrt{p})(a_j - b_j\sqrt{p}) < 0,$$

то прямоугольник с отношением сторон z *можно разрезать на прямоугольники с отношениями сторон* $x_1, \ldots, x_n$ *тогда и только тогда, когда* $z \in \{e + f\sqrt{p} > 0 \mid e, f \in \mathbb{Q}\}$;

2) *если для всех* i $(1 \leqslant i \leqslant n)$ *выполнено неравенство* $a_i - b_i\sqrt{p} > 0$, *то прямоугольник с отношением сторон* z *можно разрезать на прямоугольники с отношениями сторон* $x_1, \ldots, x_n$ *тогда и только тогда, когда*

$$z \in \left\{e + f\sqrt{p} \mid e, f \in \mathbb{Q},\ e > 0,\ \frac{|f|}{e} \leqslant \max_i \frac{|b_i|}{a_i}\right\} =: M(x_1, \ldots, x_n);$$

3) *если для всех* i $(1 \leqslant i \leqslant n)$ *выполнено неравенство* $a_i - b_i\sqrt{p} < 0$, *то прямоугольник с отношением сторон* z *можно разрезать на прямо-*

угольники с отношениями сторон $x_1, \dots, x_n$ *тогда и только тогда, когда*

$$z \in \left\{ e + f\sqrt{p} \mid e, f \in \mathbb{Q},\ f > 0,\ \frac{|e|}{f} \leqslant \max_i \frac{|a_i|}{b_i} \right\} =: N(x_1, \dots, x_n).$$

Условия второго и третьего пунктов теоремы имеют простой геометрический смысл. Изобразим число вида $z{=}e{+}f\sqrt{p}$ на координатной плоскости в виде точки (e, f). На рис. 1 *а* показан случай, когда $n = 2$, $a_1 - b_1\sqrt{p} > 0$ и $a_2 - b_2\sqrt{p} > 0$: заштрихованная область изображает множество $M(x_1, x_2)$. Это наименьший симметричный относительно оси Oe угол, который содержит все точки (a_i, b_i) $(1 \leqslant i \leqslant n)$. Аналогично на рис. 1 *б* показан случай, когда $n = 2$, $a_1 - b_1\sqrt{p} < 0$ и $a_2 - b_2\sqrt{p} < 0$: здесь изображено множество $N(x_1, x_2)$, симметричное относительно оси Of.

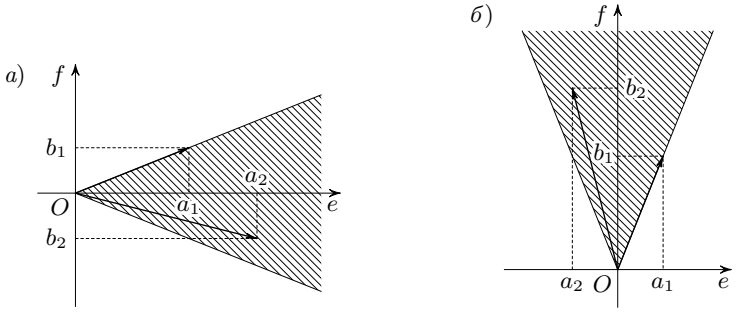


Рис. 1

Ясно, что множество $M(x_1, \dots, x_n)$ совпадает с $M(x_k)$ для некоторого $k \in \{1, \dots, n\}$. Поэтому если для всех $i \in \{1, \dots, n\}$ выполнено $a_i - b_i\sqrt{p} > 0$, то все прямоугольники, которые можно разрезать на прямоугольники с отношениями сторон $x_1, \dots, x_n$, можно разрезать и на прямоугольники с отношением сторон x_k. Аналогичное утверждение справедливо и для множества $N(x_1, \dots, x_n)$, когда для всех $i \in \{1, \dots, n\}$ выполнено $a_i - b_i\sqrt{p} < 0$.

§ 2. Обзор известных результатов

Обсуждаемой проблемой занимались многие математики (см. список литературы). Ниже приведены некоторые из полученных результатов.

Теорема 2 (Ден, 1903, см. [6]). *Если прямоугольник разрезан на квадраты (необязательно равные), то отношение его сторон рационально.*

Теорема 3 (Ден, 1903, см. теорему 1 в [7]). *Если прямоугольник с отношением сторон x можно разрезать на прямоугольники с отношениями*

сторон $x_1, x_2, \ldots, x_n$, то число x можно выразить через числа $x_1, x_2, \ldots, x_n$ с помощью сложения, вычитания, умножения и деления.

У теорем 2 и 3 существуют элементарные доказательства: для теоремы 2 оно приведено[1)] в [4], теорема 3 доказана в [3], но не сформулирована там явно. Она немедленно следует из двух результатов: теоремы о сопротивлении цепи и отношении сторон прямоугольника и леммы о сопротивлении цепи.

Теорема 4 (Лацкович, Ринн, Секереш, Фрайлинг, 1994, см. [8, 12]). *Для числа $r > 0$ следующие три условия эквивалентны*:

1) *квадрат можно разрезать на прямоугольники с отношением сторон r*;
2) *для некоторых положительных рациональных чисел c_i выполнено равенство*

$$c_1 r + \cfrac{1}{c_2 r + \cfrac{1}{c_3 r + \ldots + \cfrac{1}{c_n r}}} = 1;$$

3) *число r является корнем ненулевого многочлена с целыми коэффициентами, у которого все комплексные корни имеют положительную действительную часть.*

Частным случаем теоремы Лацковича — Ринна — Секереша — Фрайлинга является следующая теорема, элементарное доказательство которой приведено в [3].

Теорема 5. *Пусть $x = a + b\sqrt{2} > 0$, где $a, b \in \mathbb{Q}$. Тогда из прямоугольников с отношением сторон x можно составить квадрат тогда и только тогда, когда $a - b\sqrt{2} > 0$.*

В статье [7] Фрайлинг, Лацкович и Ринн свели задачу о разрезании прямоугольника на прямоугольники, подобные данному, к сложной алгебраической проблеме — они нашли алгебраический критерий возможности разрезания, правда, не дающий алгоритма её проверки. Но они полностью решили задачу для частного случая, когда отношения сторон являются квадратичными иррациональностями. Приведём эту теорему (теорема 7 в [7]) в формулировке, равносильной авторской.

Теорема 6 (Фрайлинг, Лацкович, Ринн, 1997). *Пусть $u = \alpha + \beta\sqrt{p} > 0$, где $\alpha, \beta, p \in \mathbb{Q}$ и $\beta\sqrt{p} \notin \mathbb{Q}$. Пусть также $v = \delta u + \gamma$ для некоторых рациональных γ и δ. Тогда прямоугольник с отношением сторон v можно*

[1)] См. также [2], в настоящем сборнике.

разрезать на прямоугольники с отношением сторон и в том и только том случае, когда выполнено одно из двух условий:

1) $\gamma = 0$ *и* $\delta > 0$; 2) $\alpha \neq 0$, $\dfrac{\gamma(\alpha^2 - \beta^2 p)}{\alpha} > 0$ *и* $\delta + \dfrac{\gamma}{2\alpha} \geqslant 0$.

Эта теорема равносильна случаю $n = 1$ в нашей основной теореме (равносильность доказана в § 7). Поэтому в § 6 попутно с доказательством основной теоремы мы получим и элементарное доказательство теоремы 6.

К. Китинг и Дж. Л. Кинг решили задачу, близкую к поставленной в начале статьи, — о разрезании прямоугольника на прямоугольники и так называемые антипрямоугольники (см. [9, 10]).

§ 3. Определения и обозначения

Сначала введём некоторые определения и обозначения.

Определение 1. Будем говорить, что прямоугольник P *разрезан* или *разбит* на прямоугольники $P_1, \ldots, P_n$, если $\bigcup_{i=1}^{n} P_i = P$ и любые два прямоугольника из $P_1, \ldots, P_n$ не имеют общих внутренних точек.

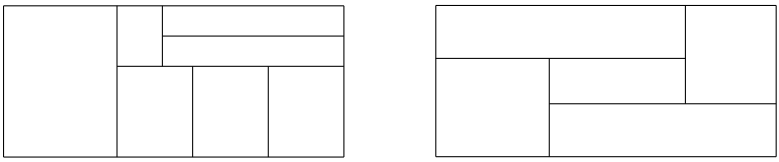
Определение 2. *Отношением сторон* прямоугольника со сторонами a и b будем называть каждое из чисел a/b и b/a.

Множества всех положительных и всех отрицательных рациональных чисел будем обозначать соответственно $\mathbb{Q}^+$ и $\mathbb{Q}^-$, положим $\mathbb{Q}^+ \cup \{0\} = \mathbb{Q}_0^+$, а также $\mathbb{R}^+ := \{r \in \mathbb{R} \mid r > 0\}$ и $\mathbb{R}^- := \{r \in \mathbb{R} \mid r < 0\}$.

Определение 3. Числа, которые можно представить в виде $x = a + b\sqrt{p}$ с рациональными a, b и p, причём p положительно и не является квадратом рационального числа, называются *квадратичными иррациональными числами* (*КИЧ*) или *квадратичными иррациональностями*. Множество всех квадратичных иррациональностей $x = a + b\sqrt{p}$ при фиксированном p будем обозначать $\mathbb{Q}[\sqrt{p}]$. Множество всех КИЧ $x = a + b\sqrt{p}$ с положительными a и b при фиксированном p будем обозначать $\mathbb{Q}^+[\sqrt{p}]$.

Определение 4. Разрезание любого прямоугольника на один прямоугольник будем называть *тривиальным*. Далее тривиальные разрезания определяются по индукции: если даны два тривиальных разрезания прямоугольников P_1 и P_2 с общей стороной, но без общих внутренних точек, то разрезание их объединения, полученное объединением данных разрезаний прямоугольников P_1 и P_2, также будем называть *тривиальным*.

На рис. 2 *а* изображено тривиальное разрезание, на рис. 2 *б* — нетривиальное. Неформально говоря, представим разрезаемый прямоугольник в виде прямоугольного листа бумаги с нарисованным на нём разбиением



а) *б*)

Рис. 2

на прямоугольники. Этот бумажный прямоугольник разрешается разрезать вдоль любого отрезка на два прямоугольника, потом производить такие операции по отдельности с каждой из получившихся частей, и так далее. Если таким образом можно реализовать исходное разбиение, то разрезание будет тривиальным.

Обозначим через $A(x_1, \ldots, x_n)$ множество всех таких чисел z, что прямоугольник с отношением сторон z можно тривиально разрезать на прямоугольники с отношениями сторон $x_1, \ldots, x_n$. Аналогично обозначим через $B(x_1, \ldots, x_n)$ множество всех таких чисел z, что прямоугольник с отношением сторон z можно разрезать на прямоугольники с отношениями сторон $x_1, \ldots, x_n$ (не обязательно тривиальным образом). Очевидно, $A(x_1, \ldots, x_n) \subset B(x_1, \ldots, x_n)$.

§ 4. Вспомогательные утверждения

Основную теорему мы докажем сначала для случая тривиальных разрезаний. В этом случае доказательство основной теоремы более наглядно, чем для произвольных разбиений. Нам понадобятся несколько вспомогательных утверждений, сформулированных в виде лемм 1–6.

Лемма 1. *Множество $A(x_1, \ldots, x_n)$ замкнуто относительно сложения, т. е. для любых чисел $a, b \in A(x_1, \ldots, x_n)$ выполнено $(a + b) \in \in A(x_1, \ldots, x_n)$.*

Доказательство. Действительно, пусть $a, b \in A(x_1, \ldots, x_n)$. Это означает, что из прямоугольников с отношениями сторон $x_1, \ldots, x_n$ мы можем тривиально составить любые прямоугольники с отношениями сторон a и b, в том числе прямоугольник со сторонами 1 и a и прямоугольник со сторонами 1 и b. Приложим эти два прямоугольника друг к другу по стороне 1. Получим прямоугольник со сторонами 1 и $a + b$. □

Лемма 2. *Множество $A(x_1, \ldots, x_n)$ замкнуто относительно операции взятия обратного числа по умножению, т. е. для любого числа $a \in A(x_1, \ldots, x_n)$ выполнено $a^{-1} \in A(x_1, \ldots, x_n)$.*

Доказательство. Действительно, если число a — отношение сторон некоторого прямоугольника, то по определению число a^{-1} также является отношением сторон того же прямоугольника. □

Лемма 3. *Множество $A(x_1, \ldots, x_n)$ замкнуто относительно умножения на положительные рациональные числа, т. е. для любого $a \in A(x_1, \ldots, x_n)$ и любого $q \in \mathbb{Q}^+$ выполнено включение $qa \in A(x_1, \ldots, x_n)$.*

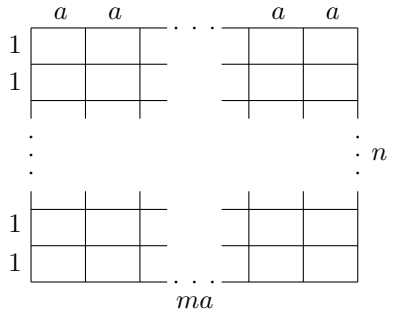


Рис. 3

Доказательство. Действительно, пусть $a \in A(x_1, \ldots, x_n)$ и $q = m/n$, где $m, n \in \mathbb{N}$. Прямоугольник со сторонами n и ma, отношение сторон которого равно $ma/n = qa$, можно тривиально разрезать на прямоугольники со сторонами 1 и a (см. рис. 3), а у этих прямоугольников, в свою очередь, существует тривиальное разбиение на прямоугольники с отношениями сторон $x_1, \ldots, x_n$, так как $a \in A(x_1, \ldots, x_n)$. □

Лемма 4. *Пусть $(a + b\sqrt{p}), (a - b\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$. Тогда из прямоугольников с отношением сторон $x = a + b\sqrt{p}$ можно тривиально сложить прямоугольник с отношением сторон $a - b\sqrt{p}$ и любой прямоугольник с рациональным положительным отношением сторон.*

Доказательство. Сначала докажем, что можно сложить прямоугольник с отношением сторон $a - b\sqrt{p}$. По лемме 2 имеем

$$\frac{a - b\sqrt{p}}{a^2 - pb^2} = \frac{1}{a + b\sqrt{p}} = \frac{1}{x} \in A(x).$$

Так как $(a^2 - pb^2) \in \mathbb{Q}^+$, по лемме 3 получаем $(a - b\sqrt{p}) \in A(x)$, что и требовалось.

Теперь докажем, что можно сложить любой прямоугольник с рациональным положительным отношением сторон. Так как

$$(a - b\sqrt{p}) + (a + b\sqrt{p}) = 2a > 0,$$

по лемме 1 прямоугольник с отношением сторон $2a$ можно тривиально сложить из двух прямоугольников с отношениями сторон $a-b\sqrt{p}$ и $a+b\sqrt{p}$. Из прямоугольников с отношением сторон $2a \in \mathbb{Q}^+$ по лемме 3 можно тривиально сложить прямоугольник с любым рациональным положительным отношением сторон. □

Лемма 5. *Пусть* $(a+b\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$ *и* $(a-b\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^-$. *Тогда из прямоугольников с отношением сторон* $x = a+b\sqrt{p}$ *можно тривиально сложить любой прямоугольник с отношением сторон* $q\sqrt{p}$ (*где* $q \in \mathbb{Q}^+$) *и прямоугольник с отношением сторон* $b\sqrt{p}-a$.

Доказательство леммы 5 аналогично доказательству леммы 4 — оставляем его читателю.

Лемма 6. *Пусть множество* P *замкнуто относительно операции сложения и операции взятия обратного по умножению. Тогда если* $x_1, \ldots, x_n \in P$, *то* $P \supset A(x_1, \ldots, x_n)$.

Доказательство. Прямоугольники с отношениями сторон $x_1, \ldots, x_n$ будем называть *базовыми*. Доказательство проведём индукцией по количеству базовых прямоугольников в разбиении.

База индукции. Из одного базового прямоугольника можно тривиально сложить лишь прямоугольники с отношениями сторон $x_1, \ldots, x_n, 1/x_1, \ldots, 1/x_n$. Из условия леммы следует, что $1/x_1, \ldots, 1/x_n \in P$. Значит, база индукции выполняется.

Шаг индукции. Пусть все прямоугольники, которые можно тривиально сложить не более чем из k прямоугольников с отношениями сторон $x_1, \ldots, x_n$, имеют отношения сторон, принадлежащие множеству P. Любой прямоугольник $ABCD$, тривиально составленный из $k+1$ базового прямоугольника, можно разрезать на два прямоугольника $ABEF$ и $ECDF$, каждый из которых разрезан не более чем на k базовых прямоугольников. По предположению индукции отношения сторон каждого из них принадлежат множеству P. Найдём отношения сторон прямоугольника $ABCD$ (см. рис. 4). Так как

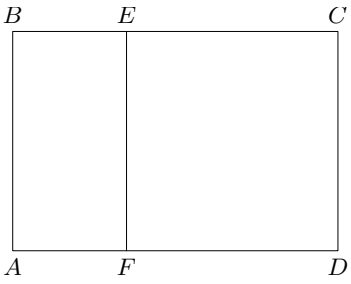


Рис. 4

$$\frac{BE}{AB}, \frac{EC}{CD} \in P,$$

получаем, что

$$\frac{BC}{AB} = \frac{BE}{AB} + \frac{EC}{CD} \in P$$

(из замкнутости P по сложению) и

$$\frac{AB}{BC} = \left(\frac{BC}{AB}\right)^{-1} \in P$$

(из замкнутости P относительно взятия обратного по умножению). $\square$

§ 5. Доказательство основной теоремы для тривиальных разрезаний

Докажем первый пункт теоремы. По условию этого пункта существуют два таких числа i и j, что $1 \leqslant i, j \leqslant n$, $a_i - b_i\sqrt{p} > 0$ и $a_j - b_j\sqrt{p} < 0$.

Докажем, что $P := \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+ \subset A(x_1, \ldots, x_n)$. Возьмём произвольное $z = (e + f\sqrt{p}) \in P$ и рассмотрим три случая.

Случай 1: $e > 0$, $f > 0$. Поскольку $(a_i + b_i\sqrt{p}), (a_i - b_i\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$ и $e \in \mathbb{Q}^+$, по лемме 4 получаем

$$e \in A(a_i + b_i\sqrt{p}) = A(x_i) \subset A(x_1, \ldots, x_n).$$

Аналогично, поскольку $(a_j + b_j\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$ и $(a_j - b_j\sqrt{p}) \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^-$ и $f \in \mathbb{Q}^+$, по лемме 5 получаем

$$f\sqrt{p} \in A(a_j + b_j\sqrt{p}) = A(x_j) \subset A(x_1, \ldots, x_n).$$

Следовательно, по лемме 1 имеем $e + f\sqrt{p} \in A(x_1, \ldots, x_n)$.

Случай 2: $e < 0$ (очевидно, что при этом $f > 0$ и $pf^2 - e^2 > 0$). Тогда $\dfrac{1}{z} = \dfrac{-e + f\sqrt{p}}{pf^2 - e^2} \in \mathbb{Q}^+[\sqrt{p}] \subset A(x_1, \ldots, x_n)$ (по случаю 1), и по лемме 2 получаем, что $z \in A(x_1, \ldots, x_n)$.

Случай 3: $f < 0$ (очевидно, что при этом $e > 0$ и $e^2 - pf^2 > 0$). Тогда $\dfrac{1}{z} = \dfrac{e - f\sqrt{p}}{e^2 - pf^2} \in \mathbb{Q}^+[\sqrt{p}] \subset A(x_1, \ldots, x_n)$ (по случаю 1), и по лемме 2 получаем, что $z \in A(x_1, \ldots, x_n)$.

Очевидно, что множество P замкнуто относительно операций сложения и взятия обратного по умножению. По лемме 6 получаем $P \supset A(x_1, \ldots, x_n)$.

Мы доказали, что $P \supset A(x_1, \ldots, x_n)$ и $P \subset A(x_1, \ldots, x_n)$, следовательно, $P = A(x_1, \ldots, x_n)$. Первый пункт теоремы доказан.

Докажем второй пункт. Покажем, что $M(x_1, \ldots, x_n) \subset A(x_1, \ldots, x_n)$. Для определённости будем считать, что $|b_1|/a_1 = \max_i(|b_i|/a_i)$, $1 \leqslant i \leqslant n$. Если $b_1 = 0$, то все числа $x_1, \ldots, x_n$ рациональны, $M(x_1, \ldots, x_n) = \mathbb{Q}^+$, а $\mathbb{Q}^+ \subset A(x_1, \ldots, x_n)$ по лемме 3. В дальнейшем будем считать, что $b_1 \neq 0$.

Пусть $z = (e + f\sqrt{p}) \in M(x_1, \ldots, x_n)$. По условию теоремы $e \in \mathbb{Q}^+$, $f \in \mathbb{Q}$ и $|f|/e \leqslant |b_1|/a_1$. Рассмотрим два случая.

Случай 1: $f \in \mathbb{Q}_0^+$. Докажем, что

$$z = e + f\sqrt{p} = \frac{f}{|b_1|}(a_1 + |b_1|\sqrt{p}) + \left(e - \frac{fa_1}{|b_1|}\right) \in A(x_1, \ldots, x_n).$$

Действительно, $(e - fa_1/|b_1|) \in \mathbb{Q}_0^+$ (так как $|f|/e \leqslant |b_1|/a_1$) и $f/|b_1| \in \mathbb{Q}_0^+$ (так как $f \in \mathbb{Q}_0^+$). По лемме 4 выполнено $(a_1 - b_1\sqrt{p}) \in A(x_1, \ldots, x_n)$, следовательно, $(a_1 + |b_1|\sqrt{p}) \in A(x_1, \ldots, x_n)$ независимо от знака b. Также по лемме 4 для любых $q \in \mathbb{Q}^+$ выполнено $q \in A(x_1, \ldots, x_n)$. Применяя леммы 1 и 3, получаем требуемое.

Случай 2: $f \in \mathbb{Q}^-$. Докажем, что

$$z = e + f\sqrt{p} = \frac{|f|}{|b_1|}(a_1 - |b_1|\sqrt{p}) + \left(e - \frac{|f|a_1}{|b_1|}\right) \in A(x_1, \ldots, x_n).$$

Действительно, $(e - |f|a_1/|b_1|) \in \mathbb{Q}_0^+$ (так как $|f|/e \leqslant |b_1|/a_1$) и $|f|/|b_1| \in \in \mathbb{Q}^+$. Применяя леммы 1 и 3, получаем требуемое.

Таким образом, $z \in A(x_1, \ldots, x_n)$, т. е. $M(x_1, \ldots, x_n) \subset A(x_1, \ldots, x_n)$.

Теперь проверим, что $M(x_1, \ldots, x_n) \supset A(x_1, \ldots, x_n)$. Очевидно, $x_1, \ldots, x_n \in M(x_1, \ldots, x_n)$, поэтому достаточно проверить замкнутость множества $M(x_1, \ldots, x_n)$ относительно операций сложения и взятия обратного по умножению и воспользоваться леммой 6.

Докажем, что $M(x_1, \ldots, x_n)$ замкнуто по сложению. Действительно, пусть $(e + f\sqrt{p}), (g + h\sqrt{p}) \in M(x_1, \ldots, x_n)$. Тогда $|f| \leqslant e|b|/a$ и $|h| \leqslant g|b|/a$. Сложив эти неравенства, получим

$$|f + h| \leqslant |f| + |h| \leqslant (e + g)\frac{|b|}{a}.$$

Кроме того, очевидно, $(e + g) \in \mathbb{Q}^+$ и $(f + h) \in \mathbb{Q}$. Следовательно,

$$\left((e + g) + (f + h)\sqrt{p}\right) \in M(x_1, \ldots, x_n).$$

Покажем, что $M(x_1, \ldots, x_n)$ замкнуто относительно операции взятия обратного по умножению. Действительно, пусть $(e + f\sqrt{p}) \in M(x_1, \ldots, x_n)$. Тогда

$$\frac{1}{e + f\sqrt{p}} = \frac{e - f\sqrt{p}}{e^2 - pf^2}.$$

Очевидно, что это число также принадлежит множеству $M(x_1, \ldots, x_n)$.

Таким образом,

$$M(x_1, \ldots, x_n) \supset A(x_1, \ldots, x_n) \quad \text{и} M(x_1, \ldots, x_n) \subset A(x_1, \ldots, x_n),$$

следовательно, $M(x_1, \ldots, x_n) = A(x_1, \ldots, x_n)$. Второй пункт нашей теоремы доказан.

Доказательство третьего пункта почти дословно повторяет доказательство второго (вместо леммы 4 нужно использовать лемму 5).

§ 6. Доказательство основной теоремы в общем случае

Приступим к доказательству основной теоремы в общем случае, т. е. без предположения тривиальности разрезаний. Для этого нам потребуется доказать ещё две леммы и дать одно определение.

Лемма 7. *Если* $x_1, \ldots, x_n \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$, *то* $B(x_1, \ldots, x_n) \subset \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$.

Доказательство. Если $x_1, \ldots, x_n \in \mathbb{Q}[\sqrt{p}]$ и $z \in B(x_1, \ldots, x_n)$, то по теореме 3 получаем, что z можно выразить через $x_1, x_2, \ldots, x_n$ с помощью сложения, вычитания, умножения и деления. Поскольку множество $\mathbb{Q}[\sqrt{p}]$ замкнуто относительно этих операций, $z \in \mathbb{Q}[\sqrt{p}]$. Число z — отношение сторон прямоугольника, поэтому $z > 0$ и $B(x_1, \ldots, x_n) \subset \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$. □

Итак, основную теорему достаточно доказать в случае $z \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$, так как это условие необходимо выполнено как при $z \in M(x_1, \ldots, x_n)$ или $z \in N(x_1, \ldots, x_n)$, так и при $z \in B(x_1, \ldots, x_n)$ (по лемме 7). Далее будем считать, что $e + f\sqrt{p} = z \in \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$.

Определение 5. Пусть $A, B, C \in \mathbb{R}$. *«Площадью»* S прямоугольника со сторонами $\alpha + \beta\sqrt{p}$ и $\gamma + \delta\sqrt{p}$, где $\alpha, \beta, \gamma, \delta, p \in \mathbb{Q}$ и $\sqrt{p} \notin \mathbb{Q}$, называется число

$$S := \alpha\gamma A + \beta\gamma B + \alpha\delta B + \beta\delta C.$$

Легко проверить, что данное определение «площади» не зависит от перестановки сторон прямоугольника.

Лемма 8. *Если прямоугольник разрезан на прямоугольники со сторонами, принадлежащими множеству* $\mathbb{Q}[\sqrt{p}]$, *то «площадь» разрезаемого прямоугольника равна сумме «площадей» прямоугольников, на которые он разрезан.*

Доказательство. Нетрудно убедиться, что сумма «площадей» двух имеющих общую сторону прямоугольников со сторонами, принадлежащими множеству $\mathbb{Q}[\sqrt{p}]$, равна «площади» их объединения. В общем случае сделаем продолжение каждого разреза, как показано на рис. 5. Очевидно, что тогда каждый прямоугольник нового разрезания будет также иметь квадратичные иррациональные стороны. Применяя утверждение предыдущего абзаца, получаем требуемое. □

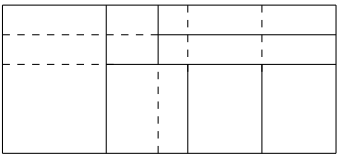


Рис. 5

Доказательство основной теоремы в случае $n = 1$. Возможность разрезаний, когда $a_1 - b_1\sqrt{p} > 0$ и $z \in M(x_1)$ или когда $a_1 - b_1\sqrt{p} < 0$ и $z \in N(x_1)$, нами уже доказана: по основной теореме для тривиальных

разрезаний, если выполнена одна из этих двух систем условий, то существует тривиальное разрезание прямоугольника с отношением сторон z на прямоугольники с отношением сторон x_1.

Остаётся доказать, что разрезание невозможно ни когда $a_1 - b_1\sqrt{p} > 0$ и $z \notin M(x_1)$, ни когда $a_1 - b_1\sqrt{p} < 0$ и $z \notin N(x_1)$. Предположим противное — пусть прямоугольник с отношением сторон z можно разрезать на прямоугольники с отношением сторон x_1 и выполнена одна из двух систем условий: либо $a_1 - b_1\sqrt{p} > 0$ и $|f|/e > |b_1|/a_1$, либо $a_1 - b_1\sqrt{p} < 0$ и $|e|/f > |a_1|/b_1$.

Будем называть разрезаемый прямоугольник с отношением сторон z *большим*, а прямоугольники с отношением сторон x_1, на которые он разрезается — *маленькими*.

Пусть большой прямоугольник имеет стороны $e + f\sqrt{p}$ и 1, т. е. $\alpha = e$, $\beta = f$, $\gamma = 1$, $\delta = 0$. По теореме о силах тока и длинах сторон и лемме о сопротивлении цепи из [3] стороны маленьких прямоугольников в этом случае должны выражаться через x_1 с помощью сложения, вычитания, умножения и деления. Тогда, поскольку множество $\mathbb{Q}[\sqrt{p}]$ замкнуто относительно этих операций, то длины сторон маленьких прямоугольников являются квадратичными иррациональностями. Таким образом, мы можем воспользоваться введённым выше определением «площади» для большого и маленьких прямоугольников.

Рассмотрим «площадь», для которой $A = f$, $B = -e$, $C = \dfrac{2fa_1^2}{b_1^2} - pf$. «Площадь» большого прямоугольника в этом случае

$$S_Б = e \cdot 1 \cdot f + f \cdot 1 \cdot (-e) + e \cdot 0 \cdot (-e) + f \cdot 0 \cdot \Big(\frac{2fa_1^2}{b_1^2} - pf\Big) = 0.$$

Теперь оценим «площадь» маленьких прямоугольников. Пусть стороны какого-то маленького прямоугольника равны $\alpha + \beta\sqrt{p}$ и $\gamma + \delta\sqrt{p}$, причём

$$\frac{\gamma + \delta\sqrt{p}}{\alpha + \beta\sqrt{p}} = a_1 + b_1\sqrt{p}.$$

Тогда $\gamma + \delta\sqrt{p} = \alpha a_1 + p\beta b_1 + \sqrt{p}(\beta a_1 + \alpha b_1)$ и выполнены равенства $\gamma = \alpha a_1 + p\beta b_1$, $\delta = \beta a_1 + \alpha b_1$. Поэтому «площадь» маленького прямоугольника записывается в виде

$$\begin{aligned} S_M &= \alpha\gamma f - \beta\gamma e - \alpha\delta e + \beta\delta\Big(\frac{2fa_1^2}{b_1^2} - pf\Big) = \alpha f(\alpha a_1 + p\beta b_1) - \\ &\quad - \beta e(\alpha a_1 + p\beta b_1) - \alpha e(\beta a_1 + \alpha b_1) + \beta\Big(\frac{2fa_1^2}{b_1^2} - pf\Big)(\beta a_1 + \alpha b_1) = \\ &= \alpha^2(fa_1 - eb_1) + 2\alpha\beta\Big(\frac{fa_1^2}{b_1} - ea_1\Big) + \beta^2\Big(\frac{2fa_1^3}{b_1^2} - pfa_1 - peb_1\Big). \end{aligned}$$

Заметим, что по предположению $|f|/e \neq |b_1|/a_1$, поэтому $fa_1 - eb_1 \neq 0$. Следовательно, выражение S_M/β^2 — квадратный трёхчлен относительно α/β. Покажем, что его дискриминант D отрицателен. Тем самым будет доказано, что при фиксированных числах a_1, b_1, e, f величина S_M либо всегда положительна, либо всегда отрицательна, а значит, сумма всех «площадей» маленьких прямоугольников не равна нулю, т. е. не равна площади большого прямоугольника. Таким образом, мы получим противоречие.

Итак,

$$\frac{D}{4} = \frac{f^2a_1^4}{b_1^2} - \frac{2fea_1^3}{b_1} + e^2a_1^2 - \frac{2f^2a_1^4}{b_1^2} + pf^2a_1^2 + pfa_1eb_1 + \frac{2fea_1^3}{b_1} - pfa_1eb_1 - pe^2b_1^2 =$$

$$= \frac{-f^2a_1^4}{b_1^2} + e^2a_1^2 + pf^2a_1^2 - pe^2b_1^2 = (a_1^2 - pb_1^2)\Big(e^2 - \frac{f^2a_1^2}{b_1^2}\Big).$$

Имеем два случая:

1) $a_1 - b_1\sqrt{p} > 0$ и $|f|/e > |b_1|/a_1 \;\Leftrightarrow\; e < a_1|f|/|b_1|$;
2) $a_1 - b_1\sqrt{p} < 0$ и $|e|/f > |a_1|/b_1 r \;\Leftrightarrow\; |e| > f|a_1|/b_1$.

В обоих этих случаях $D < 0$. □

Доказательство основной теоремы для произвольного n. Докажем первый пункт теоремы. По основной теореме для тривиальных разрезаний $B(x_1, \ldots, x_n) \supset A(x_1, \ldots, x_n) = \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$. По лемме 7 выполняется также обратное включение. Таким образом, $B(x_1, \ldots, x_n) = \mathbb{Q}[\sqrt{p}] \cap \mathbb{R}^+$, что и требуется.

Докажем второй пункт теоремы. Без ограничения общности $|b_1|/a_1 = \max_i(|b_i|/a_i)$, $1 \leqslant i \leqslant n$. Докажем, что тогда $B(x_1) = B(x_1, \ldots, x_n)$. По основной теореме для тривиальных разрезаний $x_2, \ldots, x_n \in A(x_1)$. Пусть $z \in B(x_1, \ldots, x_n)$. В разрезании прямоугольника с отношением сторон z на прямоугольники с отношениями сторон $x_1, \ldots, x_n$ заменим каждый прямоугольник с отношением сторон из множества $\{x_2, \ldots, x_n\}$, на его тривиальное разрезание на прямоугольники с отношением сторон x_1. Получим разрезание прямоугольника с отношением сторон z на прямоугольники с отношением сторон x_1. Тем самым мы доказали, что $B(x_1, \ldots, x_n) \subset B(x_1)$. Обратное включение следует из определения множества $B(x_1, \ldots, x_n)$. По основной теореме для случая $n = 1$

$$B(x_1, \ldots, x_n) = B(x_1) = \Big\{e + f\sqrt{p} \mid e \in \mathbb{Q}^+,\ f \in \mathbb{Q},\ \frac{|f|}{e} \leqslant \frac{|b_1|}{a_1}\Big\}.$$

Вспоминая, что по предположению $|b_1|/a_1 = \max_i(|b_i|/a_i)$, $1 \leqslant i \leqslant n$, получаем утверждение второго пункта основной теоремы.

Третий пункт теоремы доказывается аналогично второму. □

§ 7. Связь основной теоремы с теоремой 6

Основная теорема является обобщением теоремы 6, но на первый взгляд они выглядят совершенно разными. Покажем, что на самом деле при $n = 1$ основная теорема равносильна теореме 6.

Утверждение. *Пусть* $u = \alpha + \beta\sqrt{p}$ *и* $v = \delta u + \gamma$ (*где* $\alpha, \beta, \gamma, \delta, p \in \mathbb{Q}$ *и* $\beta\sqrt{p} \notin \mathbb{Q}$) — *положительные квадратичные иррациональности. Тогда одна из двух систем условий*:

$1^\circ)$ $\alpha - \beta\sqrt{p} > 0$ *и* $v \in M(u)$; $\quad 2^\circ)$ $\alpha - \beta\sqrt{p} < 0$ *и* $v \in N(u)$

выполнена тогда и только тогда, когда выполнена одна из двух систем условий:

$1^*)$ $\gamma = 0$ *и* $\delta > 0$; $\quad 2^*)$ $\alpha \neq 0$, $\dfrac{\gamma(\alpha^2 - \beta^2 p)}{\alpha} > 0$ *и* $\delta + \dfrac{\gamma}{2\alpha} \geqslant 0$.

Доказательство. Рассмотрим два случая: $\alpha - \beta\sqrt{p} > 0$ и $\alpha - \beta\sqrt{p} < 0$.

1) Пусть

$$\alpha - \beta\sqrt{p} > 0. \tag{1}$$

Это условие равносильно неравенству $\alpha > |\beta|\sqrt{p}$, так как $\alpha + \beta\sqrt{p} = u > 0$. Заметим также, что по условию $\beta\sqrt{p} \notin \mathbb{Q}$, т. е. $\beta \neq 0$. Далее произведём ряд равносильных преобразований, которые верны лишь при условии (1). Дано, что

$$v = \delta u + \gamma = \delta(\alpha + \beta\sqrt{p}) + \gamma = \delta\alpha + \gamma + \beta\delta\sqrt{p}.$$

Поэтому неравенство $v \in M(u)$ по основной теореме равносильно системе двух условий

$$\delta\alpha + \gamma > 0 \quad \text{и} \quad \frac{|\beta\delta|}{\delta\alpha + \gamma} \leqslant \frac{|\beta|}{\alpha}. \tag{2}$$

Если $\gamma = 0$, то система условий (2) равносильна неравенству $\delta > 0$, так как $\alpha > |\beta|\sqrt{p} > 0$. Если же $\gamma \neq 0$, то, пользуясь первым неравенством из (2) и условием $\alpha > |\beta|\sqrt{p} > 0$, второе неравенство можно преобразовать так:

$$\frac{|\beta\delta|}{\delta\alpha + \gamma} \leqslant \frac{|\beta|}{\alpha} \quad \Leftrightarrow \quad |\delta| \leqslant \delta + \frac{\gamma}{\alpha} \quad \Leftrightarrow \quad \begin{cases} -2\delta \leqslant \dfrac{\gamma}{\alpha}, \\ 0 \leqslant \dfrac{\gamma}{\alpha} \end{cases} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \begin{cases} \delta + \dfrac{\gamma}{2\alpha} \geqslant 0, \\ \gamma > 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \alpha \neq 0, \\ \delta + \dfrac{\gamma}{2\alpha} \geqslant 0, \\ \dfrac{\gamma(\alpha^2 - \beta^2 p)}{\alpha} > 0. \end{cases}$$

Получаем, что в случае (1) система условий (1°) выполнена тогда и только тогда, когда выполнена одна из двух систем условий (1^*) или (2^*).

2) Пусть

$$\alpha - \beta\sqrt{p} < 0. \tag{3}$$

Это условие равносильно неравенству $\alpha < |\beta|\sqrt{p}$, так как $\alpha + \beta\sqrt{p} = u > 0$. Заметим, что $\beta > 0$, так как иначе, очевидно, $u < 0$. Далее произведём ряд равносильных преобразований, которые верны лишь при условии (3). Включение $v \in N(u)$ по основной теореме равносильно системе двух условий:

$$\beta\delta > 0 \quad \text{и} \quad \frac{|\delta\alpha + \gamma|}{\beta\delta} \leqslant \frac{|\alpha|}{\beta}. \tag{4}$$

Если $\gamma = 0$, то система условий (4) равносильна неравенству $\delta > 0$, так как $\beta > 0$. Если же $\gamma \neq 0$, то, пользуясь первым условием из (4), а также неравенствами $\alpha < |\beta|\sqrt{p}$ и $\beta > 0$, второе условие преобразуем так:

$$\frac{|\delta\alpha + \gamma|}{\beta\delta} \leqslant \frac{|\alpha|}{\beta} \quad \Leftrightarrow \quad |\delta\alpha + \gamma| \leqslant \delta|\alpha| \quad \Leftrightarrow \quad \begin{cases} \alpha \neq 0, \\ |\delta + \dfrac{\gamma}{\alpha}| \leqslant \delta \end{cases} \quad \Leftrightarrow$$

$$\Leftrightarrow \quad \begin{cases} \alpha \neq 0 \\ \delta + \dfrac{\gamma}{2\alpha} \geqslant 0, \\ \dfrac{\gamma}{\alpha} \leqslant 0 \end{cases} \quad \Leftrightarrow \quad \begin{cases} \alpha \neq 0, \\ \delta + \dfrac{\gamma}{2\alpha} \geqslant 0, \\ \dfrac{\gamma(\alpha^2 - \beta^2 p)}{\alpha} > 0. \end{cases}$$

Получаем, что в случае (2) система условий (2°) выполнена тогда и только тогда, когда выполнена одна из двух систем условий (1*) или (2*). □

§ 8. Что дальше?

Естественным образом у читателя может возникнуть несколько вопросов, касающихся обобщения нашего результата. Ниже приведены два наиболее интересных, по нашему мнению, вопроса.

1. Какие прямоугольники можно разрезать на прямоугольники, подобные n данным, если отношения сторон данных прямоугольников — кубические иррациональности?
2. Какие прямоугольники можно разрезать на прямоугольники, подобные n данным, если отношения сторон данных прямоугольников — квадратичные иррациональности $x_1 = a_1 + b_1\sqrt{p_1}, \ldots, x_n = a_n + b_n\sqrt{p_n}$?

Автор не видит простого способа обобщить основную теорему на эти случаи. Проблема заключается в нахождении аналогов лемм 4 и 5.

В заключение приведём ещё несколько близких открытых вопросов.

1. Какие прямоугольники можно разрезать на подобные n данным?
2. Какие многоугольники можно разрезать на прямоугольники с данным отношением сторон x?

3. Какие многоугольники можно разрезать на трапеции, гомотетичные данной?

Благодарности

Автор благодарен М. Б. Скопенкову за постоянное внимание к данной работе.

Список литературы

[1] *Покровский В. Г.* О разрезаниях n-мерных параллелепипедов // Матем. заметки. 1983. Т. 33, № 2. С. 137–140.

[2] *Радзивиловский Л., Радченко Д., Танцюра М., Фещенко И.* Разрезание параллелепипеда на бруски // Математическое просвещение. Сер. 3. 2016. Вып. 20. С. 215–227

[3] *Скопенков М., Малиновская О., Дориченко С.* Собери квадрат // Квант. 2015. № 2. С. 6–11.

[4] *Скопенков М., Прасолов М., Дориченко С.* Разрезания металлического прямоугольника // Квант. 2011. № 3. С. 10–16.

[5] *Benjamini I., Schramm O.* Random walks and harmonic functions on infinite planar graphs using square tilings // Ann. Probab. 1996. V. 24, № 3. P. 1219–1238.

[6] *Dehn M.* Über die Zerlegung von Rechtecken in Rechtecke // Math. Ann. 1903. V. 57, № 3. P. 314–332.

[7] *Freiling C., Laczkovich M., Rinne D.* Rectangling a rectangle // Discr. Comp. Geom. 1997. V. 17, № 2. P. 217–225.

[8] *Freiling C., Rinne D.* Tiling a square with similar rectangles // Math. Res. Lett. 1994. V. 1. P. 547–558.

[9] *Keating K., King J. L.* Shape tiling // Electron. J. Combin. 1997. V. 4, № 2. P. R12.

[10] *Keating K., King J. L.* Signed tilings with squares // J. Comb. Theory. Ser. A. 1999. V. 85, № 1. P. 83–91.

[11] *Kenyon R.* Tilings and discrete Dirichlet problems // Israel J. Math. 1998. V. 105, № 1. P. 61–84.

[12] *Laczkovich M., Szekeres G.* Tiling of the square with similar rectangles // Discr. Comp. Geom. 1995. V. 13. P. 569–572.

[13] *Su Z., Ding R.* Tilings of orthogonal polygons with similar rectangles or triangles // J. Appl. Math. Comp. 2005. V. 17, № 1. P. 343–350.

[14] *Szegedy B.* Tilings of the square with similar right triangles // Combinatorica. 2001. V. 21, № 1. P. 139–144.

Ф. А. Шаров, НИУ ВШЭ
fedorsharov@mail.ru

По мотивам задачника

Разрезание параллелепипеда на бруски

Л. Радзивиловский, Д. Радченко, М. Танцюра, И. Фещенко

Рассмотрим множество длин всех рёбер N-мерного параллелепипеда. «Брусок» будет означать параллелепипед, у которого это множество содержит не более k элементов (определение бруска зависит от k, $1 \leqslant k \leqslant N$). Мы докажем, что параллелепипед можно разрезать на конечное количество брусков в точности тогда, когда длины его рёбер порождают линейное пространство размерности не выше k над $\mathbb{Q}$. Это обобщает известную теорему Макса Дена о разрезании прямоугольников на квадраты. Получено ещё несколько результатов о разрезаниях параллелепипедов.

§ 1. Введение

Следующий хорошо известный результат доказан Деном в 1903 г.:

Теорема 1 (Ден). *Прямоугольник можно представить как объединение неперекрывающихся квадратов в точности тогда, когда отношение его ширины к высоте рационально.*

Разумеется, вместо этого можно говорить о разрезаниях параллелограммов на ромбы.

В этой статье мы исследуем возможные многомерные обобщения теоремы 1.

Первоначальное доказательство Дена [9] было сложным. С тех пор предложено несколько доказательств этой теоремы. В 1940 г. Брукс, Смит,

Feshchenko I., Radchenko D., Radzivilovsky L., Tantsiura M. Dissecting a brick into bars // Geometriae Dedicata. 2010. V. 145, № 1. P. 159–168. Статья содержит решение задачи 13.10 из сборника «Математическое просвещение». Сер. 3. Вып. 13. М.: МЦНМО, 2009. С. 181. Перевод Б. Р. Френкина.

Стоун и Тутте [7] нашли удивительное доказательство, в котором вопрос переводится на язык электрических цепей.

В 1950-е годы Хадвигер [13] нашёл, вероятно простейший, метод, в котором с помощью функции Гамеля строятся аддитивные функции от прямоугольников [14]. Похожее доказательство было позже опубликовано В. Г. Покровским [3]. Как отметил В. Г. Болтянский [1, 2], в этих доказательствах можно обойтись без аксиомы выбора.

Другое доказательство, основанное на применении деформаций, можно найти в [4, теорема 5.6.1]. Фрейлинг и Ринне [11] опубликовали удачный обзор подобных результатов, который включает и ещё некоторые связанные с ними теоремы.

Мы воспроизведём доказательство Хадвигера — Покровского, поскольку применим похожие идеи в более общих ситуациях.

Очевидный путь обобщения теоремы Дена, отмеченный несколькими авторами, — выяснить, когда можно разрезать трёхмерный прямоугольный параллелепипед на кубы. Поскольку при этом каждая грань разрезается на квадраты, из теоремы Дена легко следует, что отношение длин любых двух рёбер параллелепипеда должно быть рационально.

Неочевидное обобщение теоремы, послужившее отправной точкой нашего исследования, — разрезание параллелепипедов на бруски.

Определение. *Трёхмерный брусок* — это параллелепипед, у которого не больше двух различных длин рёбер. Проще говоря, это коробка вида $a \times a \times b$.

Теорема 2. *В трёхмерном пространстве параллелепипед с рёбрами a, b, c можно разрезать на бруски в точности тогда, когда a, b, c связаны нетривиальным линейным соотношением с целыми коэффициентами. Другими словами, a, b, c должны порождать линейное пространство над $\mathbb{Q}$ размерности не выше* 2.

Чтобы обобщить понятие бруска на четырёхмерное пространство, возможны три способа: это будет либо параллелепипед типа $a \times a \times a \times b$ (равны три разнонаправленных ребра, всё равно какие), либо параллелепипед типа $a \times a \times b \times b$, либо оба. Приведём две теоремы, которые покажут, почему для получения результатов, аналогичных теоремам 1 и 2, нужны оба типа брусков.

Теорема 3. *В четырёхмерном пространстве можно разрезать параллелепипед вида $a \times a \times b \times b$ на параллелепипеды вида $x \times x \times x \times y$ в точности тогда, когда a, b связаны нетривиальным линейным соотношением с целыми коэффициентами.*

Теорема 4. *В четырёхмерном пространстве можно разрезать параллелепипед вида $a \times a \times a \times b$ на параллелепипеды вида $x \times x \times y \times y$ в точности тогда, когда a, b связаны нетривиальным линейным соотношением с целыми коэффициентами.*

Таким образом, теоремы 3 и 4 подсказывают, что для обобщения теоремы 2 требуется следующее

Определение. Будем называть *k-бруском* в n-мерном пространстве параллелепипед, у которого не больше k различных длин рёбер.

Теорема 5. *В n-мерном пространстве параллелепипед с длинами рёбер $a_1, a_2, \ldots, a_n$ можно разрезать на k-бруски в точности тогда, когда $a_1, a_2, \ldots, a_n$ порождают линейное пространство над $\mathbb{Q}$ размерности не выше k.*

Будем называть прямоугольник хорошим, если длина какого-либо его ребра целая. Хорошо известен следующий факт, напоминающий предыдущие теоремы:

Теорема 6. *Любой прямоугольник, который можно разрезать на хорошие, является хорошим.*

Для теоремы 6 было предложено много изящных доказательств. В [22] S. Wagon опубликовал набор из 14 доказательств, и этот набор далеко не полон, там же отмечено, что некоторые из этих доказательств можно обобщить на высшие размерности и на произвольные аддитивные подгруппы в $\mathbb{R}$, не ограничиваясь подгруппой $\mathbb{Z}$. Наш метод даёт ещё одно доказательство этой теоремы. На самом деле мы докажем следующее её обобщение. Пусть дана аддитивная подгруппа G в $\mathbb{R}$, а также число $K \leqslant N$. Назовём N-мерный параллелепипед хорошим, если G содержит длины его рёбер, идущих не менее чем в K различных направлениях.

Теорема 7. *Любой параллелепипед, который можно разрезать на хорошие, сам является хорошим.*

Теоремы 6 и 7 полезны для решения некоторых очень естественных головоломок. Однажды семилетний мальчик спросил папу, почему у него не получается заполнить коробку $6 \times 6 \times 6$ брусками $1 \times 2 \times 4$. Папа оказался математиком и для ответа на вопрос развил некую алгебраическую теорию (N. G. de Bruijn [8]). Утверждение следует из теоремы 7. В разделе 5 приведены и другие приложения наших теорем к головоломкам.

В этой статье мы докажем все вышеприведённые теоремы. Доказательства таких теорем состоят из двух частей. Надо доказать, во-первых, что при выполнении некоторого алгебраического условия параллелепипед

разрезается нужным образом и, во-вторых, что при невыполнении этого условия нужное разрезание не существует.

Первое доказывается построением нужного разрезания (его существование в некоторых случаях очевидно, но не всегда). Вторая часть доказательства состоит в построении некоторой аддитивной функции. Функции различны для различных теорем, но в их применении есть некоторые общие черты. Поэтому перед доказательствами теорем мы рассмотрим общие способы построения аддитивных функций от параллелепипедов.

§ 2. Аддитивные функции

Определение. Пусть S — множество всех параллелепипедов, грани которых параллельны граням данного параллелепипеда. Будем называть *функцию* $f\colon S \to \mathbb{R}$ *аддитивной*, если для любого $P \in S$, разрезанного на два параллелепипеда P_1 и P_2 гиперплоскостью, параллельной двум его граням, справедливо равенство $f(P) = f(P_1) + f(P_2)$.

Предложение 1. *Для любой аддитивной функции f и любого параллелепипеда P, разрезанного на n параллелепипедов $P_1, P_2, \ldots, P_n$, справедливо равенство $f(P) = f(P_1) + \ldots + f(P_n)$.*

Можно сформулировать и доказать более тонкое утверждение.

Определение. Функция $f\colon T \to \mathbb{R}$, где $T \subset S$, называется *аддитивной*, если для любого $P \in T$, разрезанного на два параллелепипеда $P_1, P_2 \in T$ гиперплоскостью, параллельной его граням, выполнено равенство $f(P) = f(P_1) + f(P_2)$.

Определение. Пусть P — параллелепипед, разрезанный на n параллелепипедов $P_1, P_2, \ldots, P_n$, и пусть координатные оси параллельны его рёбрам. Тогда каждый из параллелепипедов разрезания определяется его границами по каждой координате. Через X_k будем обозначать множество границ параллелепипедов $P_1, P_2, \ldots, P_n$ по координате x_k, а через $S' \subset S$ — множество параллелепипедов, границы которых по координате x_k принадлежат X_k при каждом k.

Предложение 1′. *Если функция $f\colon S' \to \mathbb{R}$ аддитивна, то $f(P) = f(P_1) + \ldots + f(P_n)$.*

Предложение 1′ имеет следующую мотивировку. В некоторых наших теоремах о параллелепипедах потребуется строить $\mathbb{Q}$-линейные функции над $\mathbb{R}$. При этом используется базис Гамеля и, следовательно, аксиома выбора. Было бы странно, если бы наши результаты о разрезании «коробки» на конечное количество частей зависели от аксиомы выбора, и в действи-

тельности это не имеет места. Чтобы избежать использования аксиомы выбора, можно строить $\mathbb{Q}$-линейные функции не на всём $\mathbb{R}$, а на его подходящем подпространстве, а именно конечномерном над $\mathbb{Q}$.

Чтобы не затемнять суть дела, будем использовать базисы Гамеля, но хотим отметить, что то же самое доказательство проходит и без аксиомы выбора. Те немногие читатели, которые не принимают аксиому выбора, смогут освободить от неё все наши доказательства с помощью последнего определения и предложения $1'$.

Доказательство предложения 1 можно разделить на две леммы.

Лемма 1. *Рассмотрим семейство из $n-1$ гиперплоскостей, параллельных некоторому множеству граней параллелепипеда P. Если они разрезают параллелепипед P на параллелепипеды $p_1, p_2, \ldots, p_n$, то $f(P) = f(p_1) + \ldots + f(p_n)$.*

Лемма 2. *Рассмотрим q конечных семейств параллельных гиперплоскостей, причём гиперплоскости каждого семейства соответственно параллельны некоторому множеству граней параллелепипеда P. Если они разрезают параллелепипед P на параллелепипеды $p_1, p_2, \ldots, p_n$, то $f(P) = f(p_1) + \ldots + f(p_n)$.*

Доказательство лемм. Лемма 1 непосредственно следует из определения индукцией по k. Лемма 2 доказывается индукцией по q. Базой индукции служит лемма 1.

Проведём индуктивный переход. Возьмём семейство параллельных гиперплоскостей, разрезающее параллелепипед P на $P_1, \ldots, P_m$. Каждый P_j разрезается уже лишь $q-1$ семействами гиперплоскостей на части p_i. По предположению индукции

$$f(P_1) + f(P_2) + \ldots + f(P_m) = f(p_1) + \ldots + f(p_n),$$

а по лемме 1

$$f(P) = f(P_1) + f(P_2) + \ldots + f(P_m).$$

Отсюда $f(P) = f(p_1) + \ldots + f(p_n)$. □

Предложение 1 вытекает из леммы 2 следующим образом. Продолжим все гиперплоскости граней параллелепипедов, полученных при разрезании. Они разрезают параллелепипеды $P_1, \ldots, P_m$ на меньшие части $p_1, p_2, \ldots, p_n$. Следовательно, по лемме 2

$$f(P) = f(p_1) + \ldots + f(p_n) = f(P_1) + f(P_2) + \ldots + f(P_m).$$ □

Мы видим, что это доказательство пригодно и для более сильного предложения $1'$.

Для построения аддитивных функций от параллелепипедов мы применим две конструктивные идеи.

Идея первая. Пусть $\phi_1, \phi_2, \ldots, \phi_N$ — некоторое множество аддитивных функций вещественной переменной, и пусть $a_1, a_2, \ldots, a_N$ — длины рёбер параллелепипеда P с направлениями $x_1, x_2, \ldots, x_N$. Положим

$$f(P) = \phi_1(a_1) \cdot \phi_2(a_2) \cdot \ldots \cdot \phi_N(a_N).$$

Ясно, что эта функция аддитивна, так же как любая линейная комбинация таких функций и любая полилинейная функция от $a_1, a_2, \ldots, a_N$.

Идея вторая. Параллелепипед P имеет $2N$ граней, а именно, нижнюю и верхнюю границу по каждой координате, которые и будут называться нижними и верхними гранями соответственно. Вершина параллелепипеда P будет называться чёрной, если она принадлежит чётному числу нижних граней, и белой в противном случае.

Пусть дана функция $F\colon \mathbb{R}^N \to \mathbb{R}$. Тогда можно определить следующую функцию от параллелепипедов: $f(P) =$ сумме значений функции F в чёрных вершинах параллелепипеда P минус сумма её значений в его белых вершинах.

Нетрудно видеть, что функция f аддитивна на параллелепипедах.

§ 3. Низшие размерности

Доказательство теоремы 1. Пусть w, h — ширина и высота прямоугольника.

В одну сторону доказательство очевидно: если $w/h = m/n$, где $m, n \in \mathbb{N}$, то $w/m = h/n =: a$, и прямоугольник можно замостить квадратами со стороной a.

Для доказательства в другую сторону возьмём $\mathbb{Q}$-линейную функцию $f(x)$. Тогда функция $F(\text{прямоугольник}) = w \cdot f(h) - h \cdot f(w)$ аддитивна, а на квадратах равна нулю. Если ширина и высота прямоугольника несоизмеримы, то можно построить $\mathbb{Q}$-линейную функцию f, для которой $f(w) = 0$, $f(h) = 1$ (выбрав базис в $\mathbb{R}$ над $\mathbb{Q}$, содержащий и w, и h). Тогда $F(\text{прямоугольник}) \neq 0$. Но если прямоугольник можно разрезать на квадраты, то $F(\text{прямоугольник}) = 0$ в силу аддитивности. Теорема доказана. □

Можно использовать в доказательстве и другую аддитивную функцию. Построим $\mathbb{Q}$-линейную функцию $f(x)$, положительную в точке w и отрицательную в точке h. Рассмотрим функцию

$$F(\text{прямоугольник}) = f(\text{ширина}) \cdot f(\text{высота}).$$

Она неотрицательна на любом квадрате и отрицательна на исходном прямоугольнике.

Доказательство теоремы 2 (о трёхмерных брусках и параллелепипедах). В отличие от теоремы 1, здесь доказательство в обе стороны нетривиально.

В одну сторону: пусть a, b, c линейно зависимы над $\mathbb{Z}$. Их нетривиальная линейная комбинация может иметь не более одного нулевого коэффициента. Если есть нулевой коэффициент, то имеется также положительный и отрицательный коэффициент. В этом случае условие принимает вид $ka = mb$, где $k, m \in \mathbb{N}$. Выберем два семейства гиперплоскостей — пусть одно разрезает a-рёбра на m равных частей, а другое разрезает b-рёбра на k равных частей. Тогда параллелепипед распадётся на бруски.

Если в линейной комбинации все коэффициенты ненулевые, то два из них имеют одинаковый знак, а один — противоположный. Без потери общности можно записать $ka + mb = nc$, где k, m, n — натуральные числа. Поэтому одной гиперплоскостью можно разрезать каждое ребро длины c на две части: ka/n и mb/n. Эта гиперплоскость разрежет параллелепипед на два меньших. Оба они имеют по два соизмеримых ребра разных направлений, а для этого случая задача уже решена.

В другую сторону: пусть f, g — две $\mathbb{Q}$-линейные функции над $\mathbb{R}$. Рассмотрим параллелепипед P, заданный длинами трёх рёбер: a в направлении x, b в направлении y и c в направлении z. Положим

$$F(P) = \det \begin{pmatrix} a & b & c \\ f(a) & f(b) & f(c) \\ g(a) & g(b) & g(c) \end{pmatrix}.$$

Из определения следует, что F аддитивна и равна нулю для любого бруска.

Если a, b, c линейно независимы над $\mathbb{Q}$, то можно построить такую $\mathbb{Q}$-линейную функцию f, что $f(a) = f(c) = 0$, $f(b) = 1$, и такую $\mathbb{Q}$-линейную функцию g, что $g(a) = g(b) = 0$, $g(c) = 1$. Тогда $F(P) \neq 0$ и, значит, P нельзя разрезать на бруски. □

Доказательство теоремы 3. Здесь нас интересует разрезание параллелепипеда P вида $a \times a \times b \times b$ на параллелепипеды вида $a \times a \times a \times b$.

Как мы уже знаем, если a, b соизмеримы, то параллелепипед можно разрезать даже на кубы.

Если Q — четырёхмерный параллелепипед $a \times b \times c \times d$, то положим

$$F(Q) = \det \begin{pmatrix} a & c \\ f(a) & f(c) \end{pmatrix} \cdot \det \begin{pmatrix} b & d \\ f(b) & f(d) \end{pmatrix},$$

где f — некоторая $\mathbb{Q}$-линейная функция над $\mathbb{R}$. Функция F полилинейна по a, b, c, d и, следовательно, аддитивна.

Функция F равна нулю на параллелепипедах вида $x \times x \times x \times y$, $x \times x \times y \times x$, $x \times y \times x \times x$, $y \times x \times x \times x$.

Если a, b в исходном параллелепипеде P несоизмеримы, то можно выбрать f так, что $f(a) = 0$, $f(b) = 1$. Тогда

$$F(P) = \left(\det \begin{pmatrix} a & b \\ f(a) & f(b) \end{pmatrix}\right)^2 = a^2 \neq 0.$$

Следовательно, P не разрезается на части, на которых F равна нулю. □

Доказательство теоремы 4. Здесь мы рассматриваем разрезание параллелепипедов вида $a \times a \times a \times b$ на параллелепипеды вида $x \times x \times y \times y$. Опять-таки, если a, b соизмеримы, то параллелепипед разрезается на кубы.

Пусть a, b несоизмеримы. Предположим, что разрезание возможно, и пусть f — некоторая функция вещественной переменной. Для каждого параллелепипеда P размера $x \times y \times z \times t$ положим $F(P) = f(x)f(y)f(z)f(t)$. Пусть теперь f — такая $\mathbb{Q}$-линейная функция, что $f(a) = -1, f(b) = 1$. Если R — параллелепипед с размерами $a \times a \times a \times b$, то $F(R) = -1 < 0$. Но если P имеет размеры $x \times x \times y \times y$, $x \times y \times x \times y$ или $x \times y \times y \times x$, то $F(P) = f^2(x)f^2(y) \geqslant 0$. Так как F аддитивна, R нельзя разрезать на части, на которых F неотрицательна. □

§ 4. Положительный базис и высшие размерности

Теорема о положительном базисе. *Рассмотрим*[2] *векторы* $v_1, \ldots, v_n \in \mathbb{Q}^k$, *принадлежащие открытому полупространству* H. *Тогда существует базис из* k *векторов* $e_1, \ldots, e_k \in \mathbb{Q}^k$ *того же полупространства, через который векторы* $v_1, \ldots, v_n$ *выражаются с неотрицательными коэффициентами.*

Следствие. *Если положительные вещественные числа* $x_1, \ldots, x_n$ *порождают* k*-мерное линейное пространство над* $\mathbb{Q}$, *то найдутся такие положительные числа* $e_1, \ldots, e_k$, *что* $x_1, \ldots, x_n$ — *их линейные комбинации с неотрицательными рациональными коэффициентами (и даже с неотрицательными целыми коэффициентами).*

Это следствие — в действительности частный случай теоремы о положительном базисе: положительные элементы пространства, натянутого на $x_1, \ldots, x_n$, составляют полупространство, и потому существует базис $e_1, \ldots, e_k$, через который можно выразить $x_1, \ldots, x_n$ с неотрицательными коэффициентами. Чтобы обеспечить их целочисленность, достаточно раз-

[2] В [10] дано другое доказательство этой теоремы.

делить каждый элемент базиса на наименьший общий знаменатель всех коэффициентов в этих выражениях.

Теорема о положительном базисе (точнее, её следствие) ниже применяется для доказательства теоремы 5 в одну сторону.

Доказательство теоремы о положительном базисе. Выберем базис $\{f_1, \ldots, f_{k-1}\}$ в граничной гиперплоскости B полупространства H и положим $f_k = -f_1 - \ldots - f_{k-1}$. Произвольный вектор u из B можно представить как $\alpha_1 f_1 + \ldots + \alpha_{k-1} f_{k-1}$. Без ограничения общности пусть α_1 — наименьший из этих коэффициентов. Если $\alpha_1 < 0$, то $u = -\alpha_1 f_k + (\alpha_2 - \alpha_1) f_2 + \ldots$, и в этом представлении все коэффициенты неотрицательны. В любом случае вектор из B оказывается неотрицательной линейной комбинацией векторов $f_1, \ldots, f_{k-1}, f_k$. Если теперь h — произвольный вектор из H, то $h, f_1, \ldots, f_{k-1}$ — базис в $\mathbb{Q}^k$, а выпуклая оболочка векторов $h, f_1, \ldots, f_k$ совпадает с $H \cup B$.

Выберем $\varepsilon > 0$ и положим $e_j = f_j + \varepsilon h$ $(j = 1, \ldots, k)$. Заметим, что $h = (e_1 + \ldots + e_k)/(k\varepsilon)$, т. е. h принадлежит выпуклой оболочке векторов $e_1, \ldots, e_k$, а при малом ε они соответственно близки к $f_1, \ldots, f_k$. Поэтому выпуклая оболочка векторов $e_1, \ldots, e_k$ близка к выпуклой оболочке векторов $h, f_1, \ldots, f_k$, т. е. заполняет почти всё H. При достаточно малом ε в неё попадают данные векторы $v_1, \ldots, v_n$.

Теперь достаточно показать, что $\{e_1, \ldots, e_k\}$ — базис в $\mathbb{Q}^k$. Пусть

$$\sum_{j=1}^{k} \beta_j e_j = 0.$$

Тогда

$$\sum_{j=1}^{k-1} (\beta_j - \beta_k) f_j + \varepsilon \sum_{j=1}^{k} \beta_j h = 0.$$

Следовательно, все β_j равны между собой, их сумма равна нулю и потому они сами равны нулю, что и требовалось. □

Доказательство теоремы 5. Нужно разрезать на k-бруски n-мерный параллелепипед, рёбра которого, идущие в n различных направлениях $1, \ldots, n$, равны соответственно $a_1, a_2, \ldots, a_n$.

Предположим, что $a_1, a_2, \ldots, a_n$ порождают линейное пространство над $\mathbb{Q}$ размерности не выше k. Тогда по следствию из теоремы о положительном базисе существуют положительные $e_1, e_2, \ldots, e_k$, такие что $a_1, a_2, \ldots, a_n$ линейно выражаются через них с неотрицательными целыми коэффициентами.

Следовательно, можно построить такие n семейств параллельных гиперплоскостей, что каждое семейство разрезает ребро длины a_j, идущее

в направлении j, на отрезки с длинами $e_1, e_2, \ldots, e_k$. Значит, параллелепипед разрезается на параллелепипеды с рёбрами длины $e_1, e_2, \ldots, e_k$, т. е. с не более чем k различными длинами рёбер.

Пусть теперь $a_1, a_2, \ldots, a_n$ порождают линейное пространство над $\mathbb{Q}$ размерности выше k. Тогда среди них можно выбрать подмножество из $k+1$ чисел, линейно независимых над $\mathbb{Q}$. Без ограничения общности это $a_1, a_2, \ldots, a_{k+1}$. Построим $\mathbb{Q}$-линейные функции $f_1, f_2, \ldots, f_{k+1}$ из $\mathbb{R}$ в $\mathbb{R}$, такие что при $i, j \leqslant k+1$ выполнено равенство $f_i(a_j) = \delta_{i,j}$. Для любого параллелепипеда P, рёбра которого, идущие в направлениях $1, \ldots, n$, равны соответственно $l_1, l_2, \ldots, l_n$, рассмотрим $(k+1) \times (k+1)$-матрицу $M(P)$ с элементами $m_{i,j} = f_i(l_j)$. Положим $F(P) = \det(M(P)) \cdot l_{k+2} l_{k+3} \cdot \ldots \cdot l_n$. Функция F аддитивна (поскольку полилинейна по l_j), равна 0 на брусках (поскольку тогда два столбца матрицы равны) и не равна нулю на исходном параллелепипеде. Поэтому исходный параллелепипед нельзя разрезать на бруски. $\square$

§ 5. «Хорошие» прямоугольники и параллелепипеды

Применив теперь вторую идею построения аддитивных функций, докажем утверждение, чуть более общее, чем теорема 6.

Определение. Пусть G — фиксированная аддитивная подгруппа в $\mathbb{R}$ (таким образом, $\mathbb{Z}$ — лишь один из возможных примеров). Назовём n-мерный параллелепипед хорошим, если длина хотя бы одного его ребра принадлежит G.

Теорема 6′. *Параллелепипед, который можно разрезать на хорошие параллелепипеды, является хорошим.*

Доказательство теоремы 6′. Выберем в $\mathbb{R}^n$ систему координат с осями, параллельными рёбрам параллелепипеда. Подгруппу G^n и все её параллельные сдвиги (которые являются элементами факторгруппы $\mathbb{R}^n/G^n$), будем называть решётками.

Пусть F — некоторая функция $\mathbb{R}^n \to \mathbb{R}$, значения которой определены решёткой (и не зависят от сдвигов на G^n). Такая функция не единственна, она зависит от выбора функции $\mathbb{R}^n/G^n \to \mathbb{R}$. Мы выберем её позже.

Вспомним, что вершина параллелепипеда P считается чёрной, если содержится в чётном количестве нижних граней, и белой в противном случае. Определим следующую аддитивную функцию от параллелепипедов: $f(P) =$ сумма значений функции F по чёрным вершинам параллелепипеда P минус сумма её значений по белым вершинам.

Если длина некоторого ребра принадлежит G, то можно вычеркнуть все пары вершин, соединённых параллельными ему рёбрами. Следовательно,

если исходный параллелепипед разрезается на хорошие параллелепипеды, то любая функция f указанного вида равна на нём нулю.

Но если у исходного параллелепипеда нет сторон с длинами из G, то все его вершины принадлежат разным решёткам. Поэтому можно считать F равной 1 на одной из его чёрных вершин и равной 0 на всех других вершинах, чёрных и белых. Тогда f будет равна 1 на исходном параллелепипеде. Противоречие. □

Доказательство теоремы 7. Предположим[3)], что в G входят менее чем k разнонаправленных рёбер параллелепипеда, однако разрезание на хорошие параллелепипеды существует. Обозначим разнонаправленные рёбра через $a_1, a_2, \ldots, a_n$, и пусть лишь первые из них принадлежат G, так что $a_k, \ldots, a_n$ заведомо не принадлежат G.

Рассмотрим $(n - k + 1)$-мерную грань параллелепипеда с рёбрами $a_k, \ldots, a_n$. Наше разрезание порождает разрезание этой грани на хорошие параллелепипеды в смысле теоремы $6'$, откуда получаем противоречие. □

Несколько примеров

Пример 1. Семилетний Ф. В. де Брёйн спросил своего папу [8]: можно ли заполнить коробку $6 \times 6 \times 6$ плитками $1 \times 2 \times 4$?

Ответ отрицательный. Назовём параллелепипед хорошим, если длина одного из его рёбер целая и кратна четырём. Плитки $1 \times 2 \times 4$ — хорошие, коробка — плохая, и по теореме $6'$ её нельзя заполнить этими плитками.

Пример 2 (XXIV Турнир городов, весна 2003 г., младший тренировочный вариант, задача 5). Можно ли замостить доску 2003×2003 доминошками 1×2, которые разрешается располагать только горизонтально, и прямоугольниками 1×3, которые разрешается располагать только вертикально? (Две стороны доски условно считаются горизонтальными, а две другие — вертикальными.)

Ответ отрицательный. Сожмём доску и все плитки замощения по вертикали в три раза и по горизонтали в два раза. У всех плиток одна сторона будет целой, и по теореме 6 доска, если её можно замостить, должна иметь целую сторону. Но это не выполнено.

Пример 3 (XXVI Турнир городов, осень 2004 г., старший основной вариант, задача 5). Пусть A и B — два прямоугольника. Из прямоугольников, равных A, сложили прямоугольник, подобный B. Докажите, что из прямоугольников, равных B, можно сложить прямоугольник, подобный A.

[3)] В [10] приведено и другое доказательство этой теоремы.

Доказательство. Предположим, что A имеет несоизмеримые стороны w и h. Прямоугольник B', подобный B, можно составить из прямоугольников, подобных A. По теореме 6 одна из сторон прямоугольника B' равна nw, где n целое. Точно так же у него есть сторона, равная mh, где m целое. Эти две стороны имеют разную длину, поскольку w, h несоизмеримы. Склеив $m \times n$ одинаково ориентированных экземпляров прямоугольника B, получим прямоугольник, подобный A. Если же стороны прямоугольника A соизмеримы, то соизмеримы и стороны любого прямоугольника, составленного из его копий, и утверждение задачи тривиально. □

Разумеется, для всех этих примеров существуют более элементарные доказательства, но они требуют некоторой изобретательности, тогда как теоремы 6 и $6'$ создают общий и простой метод.

Вывод: даже в трудной задаче искусная конструкция аддитивной функции может привести к изящному решению.

Благодарности

Авторы хотели бы поблагодарить профессора Джона Джейна, руководителя 15-й Международной математической олимпиады студентов университетов (IMC 15), во время которой возник замысел этой статьи. Первый из авторов благодарен Израильскому математическому союзу за спонсирование участия Израиля в этом мероприятии. Авторы также благодарны Егору Шелухину и Асафу Лави за проверку текста и ценные замечания.

Список литературы

[1] *Болтянский В. Г.* Равновеликие и равносоставленные фигуры. М.: Гостехиздат, 1956.

[2] *Болтянский В. Г.* Третья проблема Гильберта. М.: Наука, 1977.

[3] *Покровский В. Г.* О разрезаниях n-мерных параллелепипедов // Матем. заметки. Т. 33, № 2. 1983. С. 273–280.

[4] *Прасолов В. В.* Задачи и теоремы линейной алгебры. М.: МЦНМО, 2015.

[5] *Barnes F. W.* Algebraic theory of brick packing I // Discrete Math. 1982. V. 42, № 1. P. 7–26.

[6] *Barnes F. W.* Algebraic theory of brick packing II // Discrete Math. 1982. V. 42, № 2. P. 129–144.

[7] *Brooks R. L., Smith C. A. B., Stone A. H., Tutte W. T.* The dissection of rectangles into squares // Duke Math. J. 1940. V. 7, № 1. P. 312–340.

[8] *de Bruijn N. G.* Filling boxes with bricks // Amer. Math. Monthly. 1969. V. 76. P. 37–40.

[9] *Dehn M.* Über die Zerlegung von Rechtecken in Rechtecke // Math. Ann. 1903. V. 57, № 3. P. 314–332.

[10] *Feshchenko I., Radchenko D., Radzivilovsky L., Tantsiura M.* Dissecting a brick into bars // Geometriae Dedicata. 2010. V. 145, № 1. P. 159–168. http://arxiv.org/abs/0809.1883.

[11] *Freiling C., Rinne D.* Tiling a square with similar rectangles // Math. Res. Lett. 1994. V. 1. P. 547–558.

[12] *Hadwiger H.* Ergänzungsgleichheit k-dimensionaler Polyeder // Math. Z. 1952. V. 55, № 3. P. 292–298.

[13] *Hadwiger H.* Vorlesungen über Inhalt, Oberfläche und Isoperimetrie. Berlin: Springer-Verlag, 1957.

[14] *Hamel G.* Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung: $f(x+y) = f(x) + f(y)$ // Math. Ann. 1905. V. 60. P. 459–62.

[15] *Hüngerbuhler N., Nüsken M.* Delian metamorphoses // Elem. Math. 2006. V. 61. P. 1–19.

[16] *Kazarinoff N. D., Weitzenkamp R.* Squaring rectangles and squares // Amer. Math. Monthly. 1973. V. 80. P. 877–888.

[17] *Keating K.* Signed shape tilings of squares // Discrete Math. 2000. V. 215, № 1–3. P. 133–145. arXiv:math/9809118v2 [math.CO].

[18] *Keating K., King J.* Signed tilings with squares // J. Combin. Theory. Ser. A. 1999. V. 85, iss. 1. P. 83–91.

[19] *Kenyon R.* Tiling with square and square-tileable surfaces // Ecole Normale Superieure de Lyon. 1993. V. 119. P. 1–26.

[20] *King J. L. F.* Tiling stuff, an online collection of several brick-tiling articles: http://www.math.ufl.edu/~squash.

[21] *Laczkowich M., Szekeres G.* Tilings of the square with similar rectangles // Discrete Comput. Geom. 1995. V. 13. P. 569–572.

[22] *Wagon S.* Fourteen proofs of a result about tiling a rectangle // Amer. Math. Monthly. 1987. V. 94, № 7. P. 601–617.

Л. Радзивиловский, Тель-Авивский университет (Израиль)
levr78@hotmail.com

Д. Радченко, Математический институт Макса Планка (Бонн, Германия)
danrad@mpim-bonn.mpg.de

М. Танцюра, Институт математики НАН Украины
mtan@meta.ua

И. Фещенко, Киевский национальный университет им. Тараса Шевченко
ivanmath007@gmail.com

Графы с большим хроматическим числом и большим обхватом

А. М. Райгородский

§ 1. Введение

В этой статье мы поговорим об одном классическом утверждении из теории графов и о той замечательной науке, которая из него выросла. Напомним, что *хроматическим числом* графа $G = G(V, E)$ с множеством вершин V и множеством рёбер E называется величина $\chi(G)$, равная наименьшему числу цветов, в которые можно так покрасить все вершины графа, чтобы концы любого ребра имели разные цвета. Напомним также, что *обхват* графа G — это длина кратчайшего цикла в нем, обозначаемая $g(G)$ от английского слова 'girth' — «обхват».

Понятно, что если в графе есть полные подграфы, то хроматическое число графа не меньше числа вершин в каждом из них. И разумеется, чем бóльшие полные подграфы есть в графе, тем больше его хроматическое число. Интуитивно кажется довольно естественным сделать обратное предположение: если в графе нет больших полных подграфов, то он красится в малое число цветов.

Однако еще в конце 40-х годов XX века Татт и Зыков (см. [3, 17]) независимо построили для каждого k примеры графов, в которых не было треугольников (т. е. в них вовсе не было полных подграфов, кроме вершин и рёбер, а их обхваты были больше трех), но которые не красились в k цветов должным образом. Это нанесло серьезный удар по интуиции! И возник вопрос [1)]: а может быть, для каждого k и каждого l существуют графы G, у которых $\chi(G) > k$ и при этом $g(G) > l$? Это опровергло бы интуитивные

Настоящая работа выполнена при финансовой поддержке гранта РФФИ № 15–01–03530 и гранта НШ-2964.2014.1 поддержки ведущих научных школ.

[1)] В «Математическом просвещении», вып. 17, с. 197 опубликована задача 17.11 на эту тему: существует ли граф с хроматическим числом, бóльшим чем 2013, все циклы которого имеют длину больше 2013?

представления окончательно. Придумать конструкции, впрочем, не удавалось. И тем не менее в 1958 году выдающийся венгерский математик Пол Эрдёш дал утвердительный ответ (см. [18]) на поставленный вопрос! Только действовал он с помощью *вероятностного* метода. Что это значит? В следующем разделе мы докажем теорему Эрдёша, и все станет ясно.

§ 2. Доказательство теоремы Эрдёша

Зафиксируем произвольные числа k и l. Наша цель — показать, что существует граф (возможно, с очень большим количеством вершин n), у которого $\chi(G) > k$ и $g(G) > l$.

Пусть дано число $n \in \mathbb{N}$ и множество Ω_n всех графов на вершинах $V_n = \{1, \ldots, n\}$ (без петель, кратных рёбер и ориентации). Разумеется, $|\Omega_n| = 2^{C_n^2}$: любое ребро полного графа может либо присутствовать, либо отсутствовать в данном графе, а всего у полного графа C_n^2 рёбер. Возьмем число $p_n = n^{1/(2l)-1} \in (0, 1)$. Выберем *случайный* граф из множества Ω_n, полагая вероятность $\mathbb{P}(G)$ графа $G = (V_n, E)$ равной $p_n^{|E|}(1 - p_n)^{C_n^2 - |E|}$. Иными словами, граф G образуется в результате схемы испытаний Бернулли: мы C_n^2 раз бросаем монетку со смещенным центром тяжести и вероятностью «решки» p_n и всякий раз проводим очередное ребро, если монетка падает решкой кверху, и не проводим это ребро в противном случае.

В терминах теории вероятностей (см. [2, 9]) возникает *вероятностное пространство*, в котором графы суть элементарные исходы, а события — любые множества графов. Естественно, если A — событие, то

$$\mathbb{P}(A) = \sum_{G \in A} \mathbb{P}(G).$$

И, конечно, если мы знаем, что $\mathbb{P}(A) > 0$, то мы можем определенно утверждать, что множество A не пусто.

Обозначим, стало быть, через $A_n \subset \Omega_n$ событие, состоящее из графов G, у которых одновременно $\chi(G) > k$ и $g(G) > l$. Мы покажем, что $\mathbb{P}(A_n) > 0$ при достаточно больших n, и, тем самым, теорема Эрдёша будет обоснована. Замечательно то, что по ходу дела мы увидим, насколько важно было правильно подобрать «веса», т. е. вероятности рёбер p_n.

Для каждого n введем *случайную величину* $X_n \colon \Omega_n \to \mathbb{R}$, которая на данном графе $G \in \Omega_n$ принимает значение, равное количеству простых (т. е. без повторений вершин) циклов длины $\leqslant l$ в графе G. Это своего рода «число вредоносных циклов» — тех циклов, которых нет в графах из A_n.

Напомним, что *математическое ожидание* случайной величины ξ на конечном пространстве Ω — это среднее значение функции ξ, выражаемое

формулой

$$\mathbb{M}\xi = \sum_{\omega \in \Omega} \xi(\omega)\mathbb{P}(\omega). \tag{1}$$

Из этой формулы мгновенно следует *линейность математического ожидания*, а именно, для любых случайных величин ξ_1, ξ_2 и любых чисел c_1, c_2 верно равенство

$$\mathbb{M}(c_1\xi_1 + c_2\xi_2) = c_1\mathbb{M}\xi_1 + c_2\mathbb{M}\xi_2.$$

С другой стороны, из этой же формулы (1) следует, очевидно, и то, что если $\xi\colon \Omega \to \{y_1, \dots, y_k\}$, то

$$\mathbb{M}\xi = \sum_{i=1}^{k} y_i\mathbb{P}(\xi = y_i). \tag{2}$$

Вернемся к нашему пространству Ω_n и нашей случайной величине X_n. Можно представить X_n в следующем виде:

$$X_n = \sum_{r=3}^{l} X_{n,r},$$

где $X_{n,r}(G)$ — число простых циклов длины r в графе G. Далее, можно занумеровать все простые циклы C_i на r вершинах в *полном* графе K_n числами i от 1 до $C_n^r \cdot \dfrac{(r-1)!}{2}$, ведь C_n^r — это число способов выбрать вершины цикла, а $\dfrac{(r-1)!}{2}$ — это число способов построить цикл на данных r вершинах. Тогда, в свою очередь,

$$X_{n,r} = \sum_{i=1}^{C_n^r \frac{(r-1)!}{2}} X_{n,r,i}, \quad \text{где } X_{n,r,i} = \begin{cases} 1, & \text{если } C_i \subset G, \\ 0, & \text{если } C_i \not\subset G. \end{cases}$$

За счет линейности математического ожидания, а также формулы (2) получаем

$$\mathbb{M}X_n = \sum_{r=3}^{l} \sum_{i=1}^{C_n^r \frac{(r-1)!}{2}} \mathbb{P}(C_i \subset G) = \sum_{r=3}^{l} \sum_{i=1}^{C_n^r \frac{(r-1)!}{2}} p_n^r =$$

$$= \sum_{r=3}^{l} C_n^r \frac{(r-1)!}{2} p_n^r \leqslant \sum_{r=3}^{l} \frac{n^r}{r!} \frac{(r-1)!}{2} p_n^r < \sum_{r=3}^{l} n^r p_n^r.$$

У нас $np_n = n^{1/(2l)}$, откуда $\mathbb{M}X_n < l(np_n)^l = l\sqrt{n}$.

Напомним, что *неравенство Маркова* в теории вероятностей утверждает следующее (см. [9]): *Пусть ξ — случайная величина, принимающая*

только неотрицательные значения, и a *— положительное число*; *тогда*

$$\mathbb{P}(\xi > a) \leqslant \frac{\mathbb{M}\xi}{a}.$$

В нашем случае неравенство Маркова влечет оценку

$$\mathbb{P}\Big(X_n > \frac{n}{2}\Big) \leqslant \frac{\mathbb{M}X_n}{n/2} < \frac{2l}{\sqrt{n}} \to 0, \quad n \to \infty.$$

Значит, найдется такое n_1, что $\mathbb{P}(X_n > n/2) < 1/2$ при всех $n \geqslant n_1$.

Обозначая через $A_{n,1}$ событие $\{X_n \leqslant n/2\}$, получаем при $n \geqslant n_1$

$$\mathbb{P}(A_{n,1}) > \frac{1}{2}. \tag{3}$$

Иными словами, есть «много» (в смысле вероятности) графов, у каждого из которых «мало» (не больше $n/2$) вредоносных циклов.

Теперь займемся хроматическим числом. Нам нужно оценить его снизу, и мы ранее предлагали оценку числом вершин в самом большом полном подграфе (это число принято обозначать $\omega(G)$). Однако ясно, что в данной ситуации такая оценка не пригодится, ведь мы как раз стремимся избавиться и от полных подграфов, и даже от коротких циклов. Что ж, попробуем зайти с совершенно другой стороны. Напомним, что *независимым множеством* называется любое такое множество вершин графа G, что никакие два его элемента не соединены ребром. Соответственно, определим $\alpha(G)$ (*число независимости*) как мощность самого большого независимого множества вершин. Понятно, что в некотором смысле величины $\omega(G)$ и $\alpha(G)$ двойственны друг другу (отсюда и обозначения: α — первая буква греческого алфавита, ω — последняя). Оказывается,

$$\chi(G) \geqslant \frac{|V(G)|}{\alpha(G)}.$$

В сущности, это очевидно. Каждый цвет в правильной раскраске графа — это независимое множество, ведь не должно быть рёбер с концами одного цвета. Поэтому даже при правильной покраске максимальными по мощности независимыми множествами (которая не обязательно существует) потребуется не менее $\dfrac{|V(G)|}{\alpha(G)}$ цветов.

Пусть

$$m_n = \left\lceil \frac{3 \ln n}{p_n} \right\rceil.$$

Разумеется, $m_n \to \infty$ при $n \to \infty$. Рассмотрим случайную величину Y_n, равную числу независимых множеств мощности m_n в случайном графе.

Заметим, что $\mathbb{P}(\alpha(G)<m_n)=\mathbb{P}(Y_n=0)$, откуда за счет неравенства Маркова получаем $\mathbb{P}(\alpha(G) < m_n) = 1 - \mathbb{P}(Y_n \geqslant 1) \geqslant 1 - \mathbb{M}Y_n$.

Из линейности математического ожидания следует равенство

$$\mathbb{M}Y_n = C_n^{m_n}(1-p_n)^{C_{m_n}^2}.$$

Действительно, пусть $Y_{n,i}=1$, если i-е множество мощности m_n независимо, и равно 0 в противном случае. Тогда $\mathbb{M}Y_{n,i} = (1-p_n)^{C_{m_n}^2}$ и $Y_n = \sum_{i=1}^{C_n^{m_n}} Y_{n,i}$.

Далее,

$$\mathbb{M}Y_n \leqslant n^{m_n}\exp\Big(-(1+o(1))\frac{p_n m_n^2}{2}\Big) = \exp\Big(m_n\ln n - (1+o(1))\frac{p_n m_n^2}{2}\Big) =$$

$$= \exp\Big(\frac{(1+o(1))3\ln^2 n}{p_n} - \frac{(1+o(1))4{,}5\ln^2 n}{p_n}\Big) \to 0, \quad n\to\infty.$$

Значит, $\mathbb{P}(\alpha(G) < m_n) \to 1$, $n\to\infty$, т. е. найдется n_2, начиная с которого

$$\mathbb{P}(A_{n,2}) > \frac{1}{2}, \tag{4}$$

где $A_{n,2} = \{\alpha(G) < m_n\}$.

Из (3) и (4) следует, что существуют графы на $n \geqslant \max\{n_1, n_2\}$ вершинах, которые принадлежат пересечению событий $A_{n,1}, A_{n,2}$. Возьмем любой такой граф G. У него не больше $n/2$ вредоносных циклов и $\alpha < m_n$. Удалим из графа G по одной вершине из каждого вредоносного цикла. Получится новый граф G'. У этого графа число вершин не меньше $n/2$, нет вредоносных циклов (т. е. $g(G') > l$) и $\alpha < m_n$ (число независимости при удалении вершин вырасти, очевидно, не может). Значит,

$$\chi(G') \geqslant \frac{n/2}{\alpha(G')} > \frac{n/2}{m_n} = (1+o(1))\frac{np_n}{6\ln n} = (1+o(1))\frac{n^{1/2l}}{6\ln n}.$$

Сколь бы большими ни были числа l и k, отыщется такое n_3, что при $n > n_3$

$$(1+o(1))\frac{n^{1/2l}}{6\ln n} > k.$$

Таким образом, при $n \geqslant \max\{n_1, n_2, n_3\}$ граф G' и есть искомый граф. Теорема доказана.

§ 3. Вокруг теоремы Эрдёша

Теорема Эрдёша, как мы видели, носит неконструктивный характер. С одной стороны, она показывает, тем самым, «силу и славу» вероятностного метода. С другой стороны, перебор, который она сулит, необозрим,

и задача отыскания явных конструкций остается актуальной. Первым эту задачу решил Ловас в 1968 году (см. [20]). Его конструкция исключительно нетривиальна, и мы не будем приводить ее в настоящей статье.

Скажем несколько слов об одной замечательной конструкции графа с большим хроматическим числом и без треугольников — конструкции, отличной от ранних конструкций Татта и Зыкова. А именно, пусть даны натуральные числа n и $r \leqslant n/2$. Обозначим через $KG_{n,r}$ граф, у которого вершины — все возможные r-элементные подмножества множества $\{1, \ldots, n\}$, а рёбра — пары вершин, пересечение которых (как подмножеств) пусто. Этот граф называется *кнезеровским*. Очевидно, что если $r > n/3$, то в множестве $\{1, \ldots, n\}$ нет трех r-элементных подмножеств, которые бы попарно не пересекались. Это значит, что в таком кнезеровском графе нет треугольников. А что с его хроматическим числом?

Нетрудно заметить, что $\chi(KG_{n,r}) \leqslant n$. В самом деле, мы можем покрасить в i-й цвет все вершины, которые содержат (как подмножества) элемент i. Конечно, такие цвета пересекаются, но, удаляя пересечения, мы и получим нужную раскраску, ведь если множества попарно пересекаются по данному элементу, то между ними заведомо нет рёбер. Доказанную оценку можно слегка улучшить: верно, что $\chi(KG_{n,r}) \leqslant n - 2r + 2$. Действительно, рассмотрим только цвета, порожденные элементами $1, \ldots, n - 2r + 1$. Тогда непокрашенными останутся лишь те вершины, которые целиком содержатся в множестве $\{n - 2r + 2, \ldots, n\}$, имеющем мощность $2r - 1$. Однако сами-то вершины имеют мощности r, и, стало быть, точно пересекаются. Красим их в последний цвет с номером $n - 2r + 2$, и оценка установлена.

Знаменитая гипотеза Кнезера состояла в том, что $\chi(KG_{n,r}) = n - 2r + 2$. При $r > n/3$, т. е. в случаях, когда $g(KG_{n,r}) > 3$, совсем не ясно, почему эта гипотеза могла бы быть правильной. Мы знаем оценку

$$\chi(KG_{n,r}) \geqslant \frac{|V(KG_{n,r})|}{\alpha(KG_{n,r})} = \frac{C_n^r}{\alpha(KG_{n,r})}.$$

Но $\alpha(KG_{n,r}) \geqslant C_{n-1}^{r-1}$, так как мы уже приводили отличный пример независимого множества мощности C_{n-1}^{r-1}: это то самое множество вершин, каждая из которых содержит данный элемент i. А значит, из нашей оценки нельзя получить ничего лучшего, нежели неравенство

$$\chi(KG_{n,r}) \geqslant \frac{C_n^r}{C_{n-1}^{r-1}} = \frac{n}{r},$$

правая часть которого меньше трех. Факт, что $\alpha(KG_{n,r}) = C_{n-1}^{r-1}$, был доказан Эрдёшем, Ко и Радо в 1961 году (см. [6]). Гипотеза же Кнезера была

доказана упомянутым выше Ловасом с помощью придуманного им же топологического метода (см. [7, 8, 21]).

Теперь, зная, что $\chi(KG_{n,r}) = n - 2r + 2$, получаем, что, например, при $r = [n/3] + 1$ и достаточно больших n будет и $g(KG_{n,r}) > 3$, и $\chi(KG_{n,r}) > k$.

Еще одно важное направление исследований состоит в минимизации числа вершин графа с данными k, l. Но это предмет отдельного обсуждения, к которому в этой заметке мы не станем приступать. Зато в следующем разделе мы поговорим о варианте задачи, возникающем в связи с классической проблематикой комбинаторной геометрии. А в завершение этого раздела отметим, что с вероятностными методами в комбинаторике и со случайными графами можно более детально познакомиться по книгам [1, 4, 11, 15, 19].

§ 4. Теорема Эрдёша и дистанционные графы

Одной из самых известных и ярких проблем комбинаторной геометрии является задача Нелсона — Хадвигера о хроматическом числе пространства. А именно, речь идет об отыскании минимального числа $\chi(\mathbb{R}^n)$ цветов, в которые можно так покрасить все точки пространства $\mathbb{R}^n$, чтобы между точками одного цвета не было расстояния 1. Практически очевидно, что $\chi(\mathbb{R}^1) = 2$. Однако уже для плоскости все печально: известно лишь, что

$$4 \leqslant \chi(\mathbb{R}^2) \leqslant 7.$$

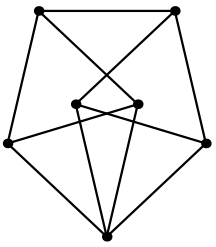


Рис. 1. Мозеровское веретено

И обе оценки очень простые! Подробности об истории задачи можно найти в обзорах и книгах [10, 12, 13, 16, 24–26], а сейчас заметим, что нижняя оценка хроматического числа плоскости получается следующим образом. Берется граф G, изображенный на рис. 1 («Мозеровское веретено»). У него рёбра — отрезки длины 1. Хроматическое число этого графа, как нетрудно видеть, равно четырем. Значит, $\chi(\mathbb{R}^2) \geqslant \chi(G) = 4$. Таким образом, мотивированным оказывается определение *дистанционного графа* в $\mathbb{R}^n$, т. е. графа, вершины которого — точки пространства, а рёбра — пары точек, отстоящих друг от друга на расстояние 1.

Тут есть одна тонкость. Иногда требуют, чтобы в дистанционном графе были проведены все рёбра — отрезки длины 1. А иногда разрешают часть рёбер не проводить. Мы будем в первом случае говорить о «полных» дистанционных графах.

При чем же здесь теорема Эрдёша? В 1976 году Эрдёш спросил: *Верно ли, что дистанционный граф на плоскости, имеющий хроматическое*

число 4, обязан содержать треугольники, подобно графу, изображенному на рис. 1? В 1979 году Уормалд дал отрицательный ответ на этот вопрос, построив пример нужного графа на 6448 вершинах. В 1996 году О'Доннелл и Хохберг уменьшили число вершин до двадцати трех! Их конструкция описана в популярной статье [5], и пока она остается рекордной. В 2000 году О'Доннелл обосновал аналог теоремы Эрдёша (см. [22,23]): *Для любого l существует дистанционный граф на плоскости с хроматическим числом 4 и обхватом, большим l.*

Еще интереснее устроена жизнь в случае, когда размерность дистанционного графа стремится к бесконечности. Здесь известно (см. [10]), что

$$(1{,}239\ldots + o(1))^n \leqslant \chi(\mathbb{R}^n) \leqslant (3 + o(1))^n,$$

т. е. по существу хроматические числа пространств растут экспоненциально. И тут имеются следующие аналоги теоремы Эрдёша, доказанные автором настоящей статьи и его учениками (см. [24]).

Теорема 1. *Существует константа $c > 1$ и такая последовательность полных дистанционных графов $G_n \subset \mathbb{R}^n$, что $g(G_n) > 3$ и $\chi(G_n) > (c + o(1))^n$.*

Теорема 2. *Для любого l существует константа $c > 1$ и такая последовательность дистанционных графов $G_n \subset \mathbb{R}^n$, что $g(G_n) > l$ и $\chi(G_n) > (c + o(1))^n$.*

Принципиальная разница между теоремами 1 и 2 в том, что доказательство теоремы 1 конструктивно и в нем дистанционные графы полны, а доказательство теоремы 2, подобно доказательству теоремы Эрдёша, вероятностное, и в нем дистанционные графы заведомо не являются полными. Отметим, что для доказательства теоремы 2 требуется существенно более трудная вероятностная техника, нежели та, которую мы использовали во втором разделе. Разумеется, здесь мы подробности не приводим.

В связи с теоремой 2 возникает вопрос: существуют ли полные дистанционные графы, обладающие характеристиками из теоремы 2? И этот вопрос исключительно труден! Сейчас удается лишь доказать следующий факт, полученный Алоном и Купавским в [14].

Теорема 3. *Для любого l существует константа c и такая последовательность полных дистанционных графов $G_n \subset \mathbb{R}^n$, что $g(G_n) > l$ и $\chi(G_n) > c\dfrac{n}{\ln n}$.*

Даже до n не дотянули! Может быть, кто-то из читателей улучшит этот результат?

Список литературы

[1] *Алон Н., Спенсер Дж.* Вероятностный метод. М.: Бином. Лаборатория знаний, 2007.

[2] *Гнеденко Б. В.* Курс теории вероятностей. М.: Физматлит, 1988.

[3] *Зыков А. А.* О некоторых свойствах линейных комплексов // Матем. сб. 1949. Т. 24, № 2. С. 163–188.

[4] *Колчин В. Ф.* Случайные графы. 2-е изд. М.: Физматлит, 2004.

[5] *Райгородский А., Рубанов О., Кошелев В.* Хроматические числа // Квант. 2008. № 3. С. 13–22.

[6] *Райгородский А. М.* Вероятность и алгебра в комбинаторике. 3-е изд. М.: МЦНМО, 2015.

[7] *Райгородский А. М.* Гипотеза Кнезера и топологические методы в комбинаторике. М.: МЦНМО, 2011.

[8] *Райгородский А. М.* Гипотеза Кнезера и топологический метод в комбинаторике // Квант. 2011. № 1. С. 7–15.

[9] *Райгородский А. М.* Комбинаторика и теория вероятностей. Долгопрудный: Интеллект, 2013.

[10] *Райгородский А. М.* Линейно-алгебраический метод в комбинаторике. 2-е изд. М.: МЦНМО, 2015.

[11] *Райгородский А. М.* Модели случайных графов. М.: МЦНМО, 2011.

[12] *Райгородский А. М.* Проблема Борсука и хроматические числа некоторых метрических пространств // УМН. 2001. Т. 56, № 1. С. 107–146.

[13] *Райгородский А. М.* Хроматические числа. 2-е изд. М.: МЦНМО, 2015.

[14] *Alon N., Kupavskii A.* Two notions of unit distance graphs // J. Comb. Theory. Ser. A. 2014. V. 125. P. 1–17.

[15] *Bollobás B.* Random Graphs. 2nd edition. Cambridge Univ. Press, 2001.

[16] *Brass P., Moser W., Pach J.* Research problems in discrete geometry. New York: Springer, 2005.

[17] *Descartes B.* A three-colour problem // Eureka. 1947. V. 9. P. 21.

[18] *Erdős P.* Graph theory and probability // Canad. J. Math. 1959. V. 11. P. 34–38.

[19] *Janson S., Łuczak T., Ruciński A.* Random graphs. New York: Wiley, 2000.

[20] *Lovász L.* On chromatic number of finite set-systems // Acta Math. Acad. Sci. Hungar. 1968. V. 19. P. 59–67.

[21] *Matoušek J.* Using the Borsuk — Ulam theorem // Universitext. Berlin: Springer, 2003.

[22] *O'Donnell P.* Arbitrary girth, 4-chromatic unit distance graphs in the plane. I. Graph description // Geombinatorics. 2000. V. 9, № 3. P. 145–152.

[23] *O'Donnell P.* Arbitrary girth, 4-chromatic unit distance graphs in the plane. II. Graph embedding // Geombinatorics. 2000. V. 9, № 4. P. 180–193.

[24] *Raigorodskii A. M.* Cliques and cycles in distance graphs and graphs of diameters // Discrete Geometry and Algebraic Combinatorics. AMS. Contemp. Math. 2014. V. 625. P. 93–109.

[25] *Raigorodskii A. M.* Coloring distance graphs and graphs of diameters // Thirty essays on geometric graph theory. Springer, 2013. P. 429–460.

[26] *Soifer A.* The mathematical coloring book. New York: Springer, 2009.

А. М. Райгородский, МГУ им. М. В. Ломоносова,
Московский физико-технический институт,
Институт математики и информатики Бурятского ГУ
mraigor@ya.ru

Задача Майхилла о синхронизации ряда стрелков

М. Д. Турбин

§ 1. Введение

В сборнике «Математическое просвещение» [3] была опубликована следующая

Задача 8.12 (задача Майхилла о стрелках). Идеальный солдат является конечным автоматом (т. е. принимает конечное число состояний и воспринимает конечное число сигналов). Солдат понимает «лево-право», воспринимает сигналы только от своих непосредственных соседей и понимает, является ли он крайним в шеренге. За один такт он обменивается сигналами с соседями. Можно ли так запрограммировать солдат, что если поставить в шеренгу солдат, находящихся в некотором одинаковом состоянии, то после того как первому [крайнему в шеренге] будет дана команда, через некоторое время все они выстрелят одновременно? (Программа не должна зависеть от длины шеренги.)

Задача вместе с её обобщением была частью проекта, предложенного на 16-й Летней конференции Турнира городов [2]. Её решила команда, состоявшая из Дмитрия Баранова и Алексея Буфетова.

Эта «задача о ряде стрелков» (предложенная Джоном Майхиллом в 1957 году) есть классическая проблема синхронизации множества автоматов. Один из её аспектов — проблема синхронизации клеточных автоматов, возникшая при наблюдении деления клеток и изучении работы сердца [1]: возбуждения клеток сердца должны быть синхронизированы; когда происходит рассинхронизация, наблюдается *фибрилляция*, которую лечат разрядом, перезапускающим сердце.

Работа поддержана грантом РФФИ № 14–01–00548.

Мы установим более сильный результат: покажем, что если команда вначале дана крайнему солдату и всего в строю n солдат, то одновременный выстрел можно осуществить через $2N - 2$ единицы времени, т. е. за время пути сигнала от одного конца шеренги до другого и обратно. Ясно, что быстрее это произойти не может. Действительно, за меньшее время первый солдат не успеет получить отклик от последнего. Если он стреляет независимо от этого, то при большей длине шеренги он выстрелит через то же самое время. Но тогда при достаточно большой длине шеренги первый солдат выстрелит до того, как последний солдат получит сигнал, а в этот момент последний солдат заведомо не выстрелит — противоречие.

Впервые алгоритм для $2N - 2$ единиц времени получен в работе [4]. Позже различным авторам удалось значительно сократить количество используемых состояний. Мы опишем новое доказательство, которое, на наш взгляд, более естественно придумывается школьником или студентом.

Вначале мы добьёмся одновременного выстрела (§ 2), затем будем оптимизировать время (§ 3). В статье изложены основные этапы доказательства для N вида $2^k + 1$ (k натуральное). Оно без труда обобщается на произвольное N. Полная версия статьи с формализованным изложением алгоритма публикуется в интернете.

§ 2. Решение задачи 8.12

Пусть в первый момент времени крайний слева стрелок посылает два сигнала, a и b. Сигнал a передаётся дальше сразу после получения, а сигнал b — с задержкой на 2 такта. Крайний справа стрелок получает инструкцию сразу после прихода сигнала отправлять его назад. Легко видеть, что сигналы встретятся в районе центра. Чтобы обработать эту ситуацию, вводим дополнительные инструкции: пусть каждый стрелок после получения сигнала b слева или сигнала a справа ждёт два такта. Если за это время стрелок получает и второй сигнал (или вообще получает их одновременно), он начинает считать себя «средним» в строю. Причём если сигналы получены одновременно, стрелок считает себя единственным средним, а а если второй сигнал получен в течение 2 тактов после первого — левым средним. В последнем случае он сообщает правому соседу об этом (специальным сигналом), и тот начинает считать себя правым средним в строю. С этого момента средние стрелки начинают считать себя крайними для соответствующих половин строя. Если средний — один, то он крайний для обеих половин. Далее, аналогичный процесс (отправка двух сигналов) осуществляется симметрично с двумя половинами строя. Когда же нужно стрелять? Введём дополнительную инструкцию: если стрелок получает

свой только что отправленный сигнал обратно через один такт, то он отправляет соседу с этой стороны сигнал «стреляй» и через один такт стреляет сам. Очевидно, что для двух и трёх солдат данная схема работает. Для остальных случаев одновременность выстрела доказывается по индукции.

Задача 8.12 решена. Отметим, что при таком алгоритме выстрел произойдёт в течение $3N$ тактов. Наша дальнейшая цель — добиться максимально быстрого выстрела, т. е. за $2N - 2$ такта.

§ 3. Как оптимизировать время

Сформулируем задачу о синхронизации стрелков в терминах автоматов. *Клеточный автомат* C есть целочисленная решётка или некоторая её часть, каждый элемент которой (*ячейка автомата*) имеет вид $M = \langle Q, q_0, N, \Delta\rangle$, где Q — множество состояний автомата; q_0 — начальное состояние автомата; $\{M, M_1, M_2, \ldots\}$ — некоторое множество клеточных автоматов, окрестность данного автомата M в решётке; Δ — множество переходов, причём $\Delta \subset Q(M) \times Q(M_1) \times Q(M_2) \times \ldots$

Таким образом, переход в новое состояние определяется состоянием всей окрестности автомата. При этом переход происходит одновременно у всех ячеек клеточного автомата. Состояние q будем представлять в виде $q = \langle f_1, f_2, \ldots, f_k\rangle$, где $f_i \in F_i$ — значения соответствующих параметров.

В этих терминах можно сформулировать задачу 8.12 следующим образом [1]:

Имеется цепь из N одинаковых автоматов. Каждый автомат имеет n возможных состояний, причём n не зависит от N. Состояние каждого автомата в следующий момент времени зависит от его состояния в текущий момент времени и состояний двух его соседей, правого и левого. В начальный момент все автоматы находятся в некотором начальном состоянии S_0. В этот момент на крайний левый автомат подаётся внешний сигнал, выводящий его из начального состояния. Существует ли такая конструкция автоматов (правила смены их состояний), что все автоматы через некоторое время перейдут одновременно в одно и то же состояние S и ни один из них не перейдёт в это состояние ни в один из предыдущих моментов?

Мы докажем, что существует решение задачи 8.12, переводящее цепочку в требуемое состояние за $2N - 2$ такта.

Проанализируем приведённое выше *решение задачи 8.12* и наметим *план доказательства* сделанного утверждения.

Пусть $N = 2^k + 1$, где k натуральное. Тогда существует центральный автомат — на расстоянии 2^{k-1} от краёв цепочки. Заметим, что если мы

будем передавать полученный сигнал дальше по цепочке в каждом такте, а после того как он дошёл до правого края — передавать его в обратную сторону, то спустя $2^k + 2^{k-1} = 3 \cdot 2^{k-1}$ тактов он окажется ровно в середине цепи, на пути назад. Также заметим, что если мы начнём передавать вместе с основным сигналом его копию, которая будет идти в 3 раза медленнее, то есть передача будет происходить в каждом третьем такте, то спустя $3 \cdot 2^{k-1}$ тактов эта копия тоже окажется ровно в середине, на пути вперёд. Таким образом, центральный автомат сможет узнать, что он центральный, спустя $3 \cdot 2^{k-1}$ тактов.

Далее он будет себя воспринимать как крайний для каждой из половин шеренги, и процесс повторится. На этом пути получается решение задачи 8.12, изложенное выше. Рассматривая сигналы различных типов, идущие с разными скоростями, их отражения и встречи, можно получить оценку на время выстрела $(2 + \varepsilon)N$, но при наивном подходе количество типов сигналов (а значит, и число состояний автомата) будет расти с уменьшением ε. Преодолеть возникшую трудность можно с помощью дополнительных сигналов, распространяющихся в обратном направлении. Пусть слева направо идёт сигнал («волна возбуждения»), и каждый автомат, до которого он дошёл, отправляет налево два специальных сигнала с разницей по времени. Когда оба эти сигнала дойдут до крайнего левого автомата, он излучает следующую волну возбуждения. Она приходит к соседнему автомату, но дальше передаётся лишь после того, как этот автомат также получит справа два специальных сигнала, и т. д. Таким способом можно породить любое количество сигналов, идущих слева направо с различными замедлениями.

Первый сигнал, возвращаясь назад после отражения от правого края, дойдёт до автомата на расстоянии 2^{k-2} от левого конца за $2^k + 3 \cdot 2^{k-2} = = 7 \cdot 2^{k-2}$ тактов. Если запустить ещё один сигнал с замедлением в 7 раз, то спустя $7 \cdot 2^{k-2}$ тактов автомат на расстоянии 2^{k-2} от левого края сможет узнать, что он центральный для левой половины автоматов. Аналогичное утверждение будет верно для сигнала с замедлением в $2^{k+1} - 1$ раз: спустя $2^{k+1} - 1 = 2N - 2 - 1$ тактов будет обнаружен автомат, находящийся на расстоянии 1 от левого края и, следовательно, центральный для цепочки из трёх самых левых автоматов.

Как только обнаружен центральный автомат во всей цепи, аналогичный процесс (излучение многих сигналов с различными замедлениями) запускается в каждой её половине, и т. д. В итоге через $2^{k+1} = 2N - 2$ тактов после начального момента все автоматы одновременно обнаружат, что они и центральные, и крайние на образовавшихся участках цепи. В этом случае они смогут одновременно перейти в состояние S, что и требуется.

Список литературы

[1] *Варшавский В. И., Поспелов Д. А.* Оркестр играет без дирижёра. М.: Наука, Главная редакция физико-математической литературы, 1984.

[2] *Иванов-Погодаев И. А., Белов-Канель А. Я.* Задачи 16-й летней Конференции Турнира Городов. Автоматы и конечно-определённые полугруппы // Математическое образование. 2004. № 3(30). С. 83–93.

[3] Математическое просвещение. Сер. 3. Вып. 8. М.: МЦНМО, 2004. С. 247.

[4] *Goto E.* A minimal time solution of the firing squad problem. Cambridge, MA: Harvard University, 1962. (Dittoed course notes for Applied Mathematics; V. 298). P. 52–59.

[5] *Moore F. R., Langdon G. G.* A generalized firing squad problem // Information and Control. 1968. V. 12, № 3. P. 212–220.

М. Д. Турбин, Московский физико-технический институт
max.turbin95@gmail.com

Нам пишут

Случайные блуждания возвращаются

А. Ю. Юрьев

Теорема Пойа о случайном блуждании по координатной сетке [1] имеет несколько различных доказательств. Оригинальное рассуждение Пойа использует производящие функции. Одно из альтернативных доказательств — «электрическое» — основывается на связи между свойствами случайных блужданий и сопротивлением электрических цепей. Авторство этого доказательства, по-видимому, не установлено (небольшую историческую справку можно найти во введении книги [2]).

В книге [2] используется утверждение о том, что вероятность возвращения для бесконечной решётки равна пределу вероятностей возвращения для возрастающей последовательности октаэдров, см. [2, стр. 68–69, п. 2.1.5], однако строгого обоснования предельного перехода авторы не дают. Аналогичное утверждение оставлено без доказательства и в статье [3].

Эта заметка устраняет пробел в обосновании предельного перехода в «электрическом» доказательстве теоремы Пойа для случайного блуждания по трёхмерной сетке, приведённом в статье [3, стр. 45]. Уточнённое доказательство также использует только элементарные методы.

Рассматривается случайное блуждание по трёхмерной целочисленной сетке. Блуждание начинается в точке $(0, 0, 0)$, и на каждом шаге к одной из координат добавляется ± 1, при этом каждое из 6 направлений перемещения выбирается с равными вероятностями.

Пусть P — вероятность того, что при случайном блуждании по трёхмерной сетке мы когда-нибудь вернёмся в начальную точку. По определению, число P равно пределу последовательности P_T, $T \to \infty$, где P_T — вероятность вернуться в начальную точку, сделав не более T шагов (этот предел

существует, так как P_T не убывает по T и ограничена сверху единицей; из этого также следует, что $P_T \leqslant P \leqslant 1$). Теорема Пойа утверждает, что $P < 1$.

Определение. Будем называть октаэдром с диагональю $2n$ подграф трёхмерной сетки с вершинами, имеющими координаты (x, y, z), причём $|x| + |y| + |z| \leqslant n$, соответственно его границей — все вершины, для которых $|x| + |y| + |z| = n$.

Рассмотрим случайное блуждание по октаэдру с диагональю $2n$. Во внутренних вершинах октаэдра этот процесс ничем не отличается от блуждания по трёхмерной сетке, а в граничных из 6 направлений доступна только часть.

Определение. Пусть $P_T^{(n)}$ — вероятность вернуться в центр октаэдра с диагональю $2n$, не попадая на его границу и сделав не более T шагов. Тогда вероятностью возврата в центр октаэдра с диагональю $2n$ до достижения его граничных точек называется величина $P^{(n)}$ — предел $P_T^{(n)}$ при $T \to \infty$.

Рассуждая по аналогии с P_T и P, получаем, что определение корректно (предел существует), и, кроме того, $P_T^{(n)} \leqslant P^{(n)} \leqslant P \leqslant 1$.

Теорема. *Верхняя оценка величины P_T имеет вид $P_T \leqslant P^{(n)}$ при $n > T$.*

Доказательство. Пусть $n > T$, тогда очевидно, что до границы октаэдра с диагональю $2n$ нельзя добраться за T или меньше шагов. Из этого следует, что $P_T = P_T^{(n)}$, так как при таких ограничениях на T блуждание по октаэдру не отличается от блуждания на всей трёхмерной сетке. Пользуясь тем фактом, что $P_T^{(n)} \leqslant P^{(n)}$, приходим к требуемой оценке: $P_T = P_T^{(n)} \leqslant P^{(n)}$. □

Перейдём к доказательству теоремы Пойа. При доказательстве мы будем ссылаться на леммы, доказанные в статье [3].

Доказательство теоремы Пойа. Возьмём произвольное $i > 0$, $n = 2^i - 1$ и октаэдр с диагональю $2n$. Пусть $C^{(n)}$ — проводимость между началом координат и границей этой фигуры. По лемме о вырезании дерева [3, стр. 45], из такой части сетки можно вырезать модифицированное троичное дерево глубины i с пересечениями рёбер на равном расстоянии от корня. Легко заметить, что проводимость дерева с такими пересечениями равна проводимости такого же дерева без пересечений.

По лемме о проводимости дерева [3, стр. 44], проводимости модифицированных троичных деревьев любой глубины больше 1. Поэтому и проводимости вырезаемых деревьев с пересечениями больше 1. По принципу монотонности [3, стр. 38] получаем, что $C^{(n)} > 1$.

Из физической интерпретации вероятности возврата [3, стр. 37] следует, что $P^{(n)} = 1 - C^{(n)}/6 < 5/6$.

Для произвольного T у нас есть верхняя оценка величины P_T: если взять достаточно большое i, а именно такое, что $2^i - 1 = n > T$, то P_T не будет превосходить $P^{(n)} < 5/6$.

Осталось выполнить предельный переход $P_T \to P$ и получить, что вероятность P возврата в начальную точку также не превосходит $5/6$. □

Указание к задаче 10 из статьи [3]. Вырезать из квадрата $2n \times 2n$ модифицированное двоичное дерево с корнем в центре квадрата (см. задачу 12; при этом допускается склейка вершин на равных расстояниях от корня).

В заключение исправим несколько неточностей в статье [3], замеченных К. Душиным и Ф. Шаровым.

Стр.	Строка	Напечатано	Должно быть
25	16 сверху	определения	иллюстрации (формальные определения даны в § 8 статьи «Сквозь сеть сопротивлений» в вып. 18 за 2014 год)
29	2 снизу	$b = (a + c + 2)/4$	$b = (a + e + 2)/4$
31	15 снизу	электрической цепи	электрической цепи с единичными проводимостями рёбер
32	6 снизу	возрастает	не убывает
34	16 снизу	$u(x) - \varepsilon(x)$	$u(x) + \varepsilon(x)$
34	13 снизу	$u(x) - \varepsilon(x) - u(y) + \varepsilon(y)$	$u(x) + \varepsilon(x) - u(y) - \varepsilon(y)$
35	7 снизу	резисторов	резисторов (т. е. резисторов ровно с одной общей вершиной, не граничной и не принадлежащей другим резисторам)
37	1 сверху	новой вершины	новой вершины (если она не граничная)
37	9 снизу	Если он начинает движение в точке a, то	Выходя из очередной вершины, он случайным образом выбирает, по какому ребру двигаться дальше, и проходит его целиком.
40	5 снизу	в качестве следствия	рассуждая аналогично доказательству
41	1 сверху	каких-либо вершин	каких-либо неграничных вершин

Список литературы

[1] *Pólya G.* Uber eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Straßennetz // Math. Ann. 1921. V. 84. P. 149–160.

[2] *Doyle P. G.*, *Snell J. L.* Random walks and electric networks // Mathematical Association of America, 1984 http://arxiv.org/abs/math/0001057v1.

[3] *Скопенков М., Смыкалов В., Устинов А.* Случайные блуждания и электрические цепи // Математическое просвещение. Сер. 3. Вып. 16. М.: МЦНМО, 2012. С. 25–47.

А. Ю. Юрьев, Независимый московский университет
alexander.yuriev@gmail.com

Непустота пересечения цепочки множеств

А. Д. Матушкин

Теорема. *Пусть $n \geqslant 2$ и $A_1, \ldots, A_n$ — подмножества конечного множества M, доля каждого из которых в M (т. е. $|A_k|/|M|$) больше, чем $1 - \frac{1}{n-1}$. Если A_k и A_{k+1} независимы (т. е. $|A_k| \cdot |A_{k+1}| = |A_k \cap A_{k+1}| \cdot |M|$) для любого $k = 1, 2, \ldots, n-1$, то $A_1 \cap A_2 \cap \ldots \cap A_n \neq \varnothing$.*

Данная теорема сформулирована в работе [1] в качестве задачи 11(b). По словам авторов, им не было известно её решение. Теорема была доказана независимо А. Ляховцом и О. Орёл.

Доказательство теоремы. При $n = 2$ утверждение теоремы тривиально. Пусть $n > 2$. Оценим мощность пересечения $A_1 \cap A_2 \cap \ldots \cap A_n$:

$$\begin{aligned} |A_1 \cap \ldots \cap A_n| &= |A_2 \cap \ldots \cap A_n| - |\overline{A}_1 \cap A_2 \cap \ldots \cap A_n| \geqslant \\ &\geqslant |A_2 \cap \ldots \cap A_n| - |\overline{A}_1 \cap A_2| \geqslant \\ &\geqslant |A_3 \cap \ldots \cap A_n| - |\overline{A}_2 \cap A_3| - |\overline{A}_1 \cap A_2| \geqslant \ldots \geqslant \\ &\geqslant |A_n| - \sum_{k=1}^{n-1} |\overline{A}_k \cap A_{k+1}|. \end{aligned}$$

Положим $x_i := \dfrac{|A_i|}{|M|}$. В силу независимости множеств A_k и A_{k+1} имеем

$$\frac{|\overline{A}_k \cap A_{k+1}|}{|M|} = (1 - x_k)x_{k+1}.$$

Следовательно,

$$\frac{|A_1 \cap \ldots \cap A_n|}{|M|} \geqslant x_n - \sum_{k=1}^{n-1} (1 - x_k)x_{k+1} =: f(x_1, \ldots, x_n).$$

Заметим, что функция $f(x_1, \ldots, x_n)$ линейна по каждому из своих аргументов x_k, $k \in \{1, \ldots, n\}$, причем коэффициент при x_k положителен: он равен

$x_2 > 0$ при $k = 1$, $x_{k+1} + x_{k-1} - 1 > 2 - \frac{2}{n-1} - 1 \geqslant 0$ при $k \in \{2, \ldots, n-1\}$ и $x_{n-1} > 0$ при $k = n$. Следовательно, на множестве

$$X = \left\{(x_1, \ldots, x_n)\colon 1 \geqslant x_i \geqslant 1 - \frac{1}{n-1}\right\}$$

функция $f(x_1, \ldots, x_n)$ возрастает по каждому из аргументов. Поэтому

$$f(x_1, \ldots, x_n) > f\Big(\frac{n-2}{n-1}, \ldots, \frac{n-2}{n-1}\Big) = \frac{n-2}{n-1} - (n-1) \cdot \frac{n-2}{(n-1)^2} = 0.$$

Получаем, что $|A_1 \cap A_2 \cap \ldots \cap A_n| > 0$. Что и требовалось доказать. □

Список литературы

[1] *Ильинский Д. Г., Райгородский А. М., Скопенков А. Б.* Независимость и доказательства существования в комбинаторике // Математическое просвещение. Сер. 3. Вып. 19. М.: МЦНМО, 2015. С. 164–177. http://arxiv.org/pdf/1411.3171.pdf.

А. Д. Матушкин, Московский физико-технический институт
1alexmatushkin1@gmail.com

Задачник

Условия задач

В этом разделе вниманию читателей предлагается подборка задач разной степени сложности, в основном трудных. Надеемся, что эти задачи окажутся интересными как для сильных школьников, интересующихся математикой, так и для студентов-математиков.

Мы обращаемся с просьбой ко всем читателям, имеющим собственные подборки таких задач, присылать их в редакцию. И, разумеется, мы с удовольствием будем публиковать свежие авторские задачи.

В скобках после условия задачи приводится фамилия автора (уточнения со стороны читателей приветствуются).

1. Частица движется по прямой линии, при этом направление движения может меняться. В каждый момент времени ускорение частицы не превосходит 1 м/с^2 по абсолютной величине. Через одну секунду после начала движения частица вернулась в начальную точку. Докажите, что её скорость через 0,5 с после начала движения не превосходит 0,25 м/с. (*А. А. Колчев*)

2. а) В клетках квадрата 5×5 стоят знаки $+$ и $-$. Можно положить пятиклеточный крест центральной клеткой на клетку квадрата и поменять все знаки, закрытые крестом, на противоположные. Можно ли с помощью таких операций все знаки поменять на противоположные? (*Р. Г. Женодаров*)

   б) Дан граф, в вершинах которого стоят знаки $+$ и $-$. Разрешается выбрать вершину и поменять знаки на противоположные в ней и в смежных вершинах. Докажите, что такими действиями можно заменить все знаки на противоположные. (*И. В. Митрофанов*)

3. Укажите на комплексной плоскости множество нулей всех многочленов с положительными коэффициентами степени не выше n. (*Р. М. Тригуб*)

4. Последовательность $\{a_n\}$ называется *линейной рекуррентой порядка* k, если для некоторых $b_1, \ldots, b_k$ при всех $n \geqslant k$ выполняется равенство $b_0 a_n + b_1 a_{n-1} + \ldots + b_k a_{n-k} = 0$. Пусть $b_0 = 1$, $b_i \in \mathbb{Z}$ при всех i. Докажите, что последовательность $\{a_n\}$ либо содержит член, имеющий не менее 2016 различных простых делителей, либо является геометрической прогрессией: $a_n = c \cdot d^n$. (*А. Я. Канель-Белов*)

5. Имеется 9 точек на плоскости: три красные, три синие, три зелёные. Никакие 4 не лежат на одном цикле (окружности или прямой). Дугами циклов будем называть дуги окружностей и отрезки прямых. Докажите, что существуют три непересекающиеся дуги цикла, на каждой из которых лежит по точке каждого цвета. (*A. Oppenheim*)

6. Пусть точка D лежит на малой оси эллипсов Штейнера треугольника ABC на расстоянии $f\sqrt{2}$ от его центра, где f — расстояние от центра описанного эллипса Штейнера до его фокуса. Докажите, что C лежит на малой оси эллипсов Штейнера треугольника ABD. (Эллипсами Штейнера называются образы вписанной и описанной окружностей правильного треугольника при аффинном преобразовании, переводящем его в данный треугольник.) (*А. А. Заславский*)

7. Имеется n монет, разложенных в несколько стопок. За одну операцию из каждой непустой стопки берётся по монете, из них формируется новая стопка.
   а) Докажите, что если $n = \binom{k}{2}$, то ситуация стабилизируется: возникнут стопки из $1, \ldots, k$ монет. Каков возможный период в общем случае?
   б) Укажите максимально возможную длину предпериода. (*Фольклор*)

8. а) Дан полный граф на n вершинах. Двое по очереди красят его рёбра. Первый красит одно ребро красным, второй красит 100 рёбер в синий цвет, и так повторяется, пока все рёбра не будут покрашены. Может ли первый при достаточно больших n добиться появления полного подграфа со 100 вершинами и всеми красными рёбрами? (*А. Я. Канель-Белов*)
   б) В красный цвет раскрашены 99 % рёбер полного графа из n вершин. Верно ли, что при достаточно большом n найдётся полный подграф из 1000 вершин с рёбрами одного цвета? (*А. Я. Канель-Белов*)
   в) Каждое k-элементное подмножество множества $\{1, \ldots, n\}$ раскрашено в один из s цветов (например, при $k = 2$ получаем раскраску рёбер полного графа на n вершинах). Докажите, что при достаточно большом n найдётся такое подмножество $U \subset \{1, \ldots, n\}$, что все его k-элементные

подмножества — одного и того же цвета, причём если x — минимальный элемент из U, то число элементов в U не меньше $x + 2s$. (*Фольклор*)

9. а) Матрица A называется *нормальной*, если $AA^* = A^*A$, где A^* — матрица, транспонированная к матрице A. Какова максимальная размерность векторного подпространства комплексных $n \times n$-матриц, все элементы которого нормальны? (*Международная студенческая олимпиада, 2015*)
б) Какова максимальная размерность подпространства попарно коммутирующих комплексных $(n \times n)$-матриц? (*Э. Б. Винберг*)

10. Дан многочлен $P(x)$ степени n с целыми коэффициентами, e — основание натуральных логарифмов. Докажите, что $\max_{x \in [0,1]} |P(x)| > e^{-n}$. (*Международная студенческая олимпиада, 2015*)

11. (Задача на исследование.) Пусть k и n — фиксированные натуральные числа. Двое играют в следующую игру. Вначале все целые числа считаются непокрашенными. На каждом ходу первый игрок выбирает непокрашенное целое число и красит его в красный цвет, а второй игрок выбирает k непокрашенных целых чисел и красит их в синий цвет. Первый стремится создать арифметическую прогрессию длины n, а второй стремится ему помешать.
а) Докажите, что при любых n и k у первого есть выигрышная стратегия (алгоритм, позволяющий выиграть при любых действиях второго).
б) Для $n = 3$ найдите минимальное число ходов, за которое первый заведомо сможет выиграть.
в) Постарайтесь оценить минимальное необходимое число ходов в общем случае. Верна ли полиномиальная или хотя бы экспоненциальная оценка? (*А. Я. Канель-Белов*)

12. а) Дан n-мерный шар единичного объёма. Пусть M_1 и M_2 — два его подмножества объёма ε. Докажите, что расстояние между ними ограничено функцией, зависящей от ε, но не от размерности.
б) Аналогичный вопрос для n-мерного куба единичного объёма. (*Н. А. Бобылев, А. Я. Канель-Белов*)

Решения задач из прошлых выпусков

В скобках после условия задачи указан её автор, а после решения — автор решения.

6.10. Условие. Для тройки прямых можно определить окружность — окружность, описанную около соответствующего треугольника. Для четвёрки прямых общего положения можно определить точку — как пересечение всех окружностей троек прямых (исходная четвёрка без одной). Для пятёрки прямых можно определить окружность — как окружность, проходящую через все точки четвёрок, и т. д. Докажите, что вся эта цепочка определений корректна. (*Фольклор*)

Решение. Переформулируем утверждение задачи следующим образом. Назовём точку пересечения двух прямых центральной точкой этих прямых, а окружность, описанную около треугольника, — центральной окружностью трёх прямых, содержащих его стороны. Тогда надо доказать следующее утверждение.

Пусть даны n прямых ($n \geqslant 4$). Если n чётно, то пусть определены центральные окружности любых $n-1$ из этих прямых. Тогда они пересекаются в одной точке (назовём эту точку центральной точкой n прямых). Если n нечётно, то пусть определены центральные точки любых $n-1$ из этих прямых. Тогда они лежат на одной окружности (назовём её центральной окружностью n прямых).

При $n=4$ утверждение задачи хорошо известно (искомая точка называется точкой Микеля данных прямых) и доказывается простым счётом углов (см. *Прасолов В. В.* Задачи по планиметрии. М.: МЦНМО, 2007. Задача 2.88. С. 40, 52–53). Сделав инверсию с центром в некоторой точке O, получим, что справедлива следующая

Лемма. *Пусть четыре окружности проходят через точку O. Тогда окружности, каждая из которых проходит через вторые точки пересечения трёх из данных окружностей, пересекаются в одной точке.*

Теперь утверждение задачи можно доказать по индукции. Пусть $n > 4$ и для меньших значений n утверждение верно.

Пусть n чётно. Обозначим данные прямые числами $1, \ldots, n$, центральную окружность прямых $2, \ldots, n$ через c_1, прямых $1, 3, \ldots, n$ — через c_2 и т. д. Аналогично центральную точку прямых $3, \ldots, n$ обозначим через A_{12} и т. д., центральную окружность прямых $4, \ldots, n$ через c_{123} и т. д., центральную точку прямых $5, \ldots, n$ через A_{1234} и т. д. (существование всех этих окружностей и точек следует из предположения индукции). Достаточно доказать, что окружности c_1, c_2, c_3, c_4 пересекаются в одной точке.

Из определения центральных окружностей и точек следует, что:

— окружности c_{123}, c_{124}, c_{134}, c_{234} проходят через точку A_{1234};
— точка A_{12} лежит на окружностях c_{123} и c_{124}, аналогичные утверждения верны для всех точек A_{ij} при $i, j = 1, \ldots, 4$;
— окружность c_1 проходит через точки A_{12}, A_{13}, A_{14} и т. д.

Поэтому искомое утверждение непосредственно следует из леммы, если в качестве O взять A_{1234}.

Пусть n нечётно. Обозначим данные прямые через $1, \ldots, n$, центральную точку прямых $2, \ldots, n$ через A_1, центральную окружность прямых $3, \ldots, n$ обозначим через c_{12}, центральную точку прямых $4, \ldots, n$ через A_{123}, центральную окружность прямых $5, \ldots, n$ через c_{1234} и т. д. (существование всех этих окружностей и точек следует из предположения индукции). Достаточно доказать, что точки A_1, A_2, A_3, A_4 лежат на одной окружности.

Из определения центральных окружностей и точек следует, что:

— точки A_{123}, A_{124}, A_{134}, A_{234} лежат на окружности c_{1234};
— окружность c_{12} проходит через точки A_{123} и A_{124}, и т. д.;
— точка A_1 лежит на окружностях c_{12}, c_{13}, c_{14}, и т. д.

Рассмотрим окружности c_{12}, c_{13}, c_{23} и c_{1234}. Они проходят через точку A_{123}, а вторыми точками их пересечения являются A_1, A_2, A_3, A_{124}, A_{134}, A_{234}. При этом описанными окружностями треугольников $A_1A_{124}A_{134}$, $A_2A_{124}A_{234}$ и $A_3A_{134}A_{234}$ будут окружности c_{14}, c_{24}, c_{34} соответственно. Эти окружности пересекаются в точке A_4, и по лемме окружность $A_1A_2A_3$ также проходит через эту точку.

Комментарий. Эта задача и её решение имеются в книге: *Шклярский Д. О., Ченцов Н. Н., Яглом И. М.* Избранные задачи и теоремы элементарной математики. Часть 2. М.: Гостехиздат, 1952. С. 40–41, 295–303.

(*А. А. Заславский*)

12.6. Условие. Докажите, что монотонная функция дифференцируема в некоторой точке. (*Теорема Лебега*)

Решение. Докажем, что монотонная функция дифференцируема почти всюду. Эта теорема обычно излагается в курсе теории меры Лебега. Приводимое доказательство не использует понятия меры множества и измеримости, но ограничивается понятием множества меры 0.

Определение. Множество $A \subset \mathbb{R}$ назовём *множеством меры* 0, если для всякого $\varepsilon > 0$ его можно покрыть счётной системой интервалов с суммой длин меньше ε.

Заметим, что класс Null множеств меры 0 замкнут относительно

i) взятия подмножеств: если $A \in \text{Null}$, $B \subset A$, то $B \in \text{Null}$ (очевидно);

ii) не более чем счётных объединений: если $A_i \in \text{Null}$ при всех $i = 1, 2, \ldots$, то и $\cup A_i \in \text{Null}$. В самом деле, при данном ε множество A_n можно покрыть интервалами с суммой длин меньше чем $\varepsilon/2^n$, тогда суммарная длина всех интервалов будет меньше ε.

Слова «почти все точки множества $E \subset \mathbb{R}$» означают «точки множества $E \setminus E_0$, где E_0 — некоторое множество меры 0». Один из популярных способов доказать, что некоторое подмножество промежутка $\Delta \subset \mathbb{R}$ не пусто — установить, что ему принадлежат почти все точки промежутка. Этого достаточно, поскольку имеет место фундаментальный факт.

Теорема 1. *Если отрезок покрыт интервалами, то сумма длин интервалов больше длины отрезка.*

Доказательство. Докажем сначала, что если отрезок покрыт n интервалами, то сумма их длин больше, чем длина отрезка. Индукция по n. База $n = 1$ очевидна. Переход от $n - 1$ к n: если отрезок $[a, b]$ покрыт n интервалами, пусть (c, d), $c < a < d$, — тот из них, что содержит точку a. Тогда отрезок $[d, b]$ покрыт $n - 1$ интервалом, сумма их длин по индукционному предположению больше чем $b - d$, а длина интервала (c, d) больше $d - a$, в сумме получается больше чем $b - a$. Индукционный переход завершён. А из бесконечного покрытия отрезка интервалами можно выбрать конечное подпокрытие. □

Докажем, что множество точек, в которых не дифференцируема данная монотонная (пусть возрастающая) функция $f(x)$, имеет меру 0. Ключевую роль играет

Теорема 2 (теорема Витали о покрытии). *Предположим, что множество $E \subset \mathbb{R}$ покрыто отрезками таким образом, что каждая точка $x \in E$ содержится в сколь угодно малом отрезке покрытия. Тогда из отрезков*

покрытия можно выбрать непересекающуюся конечную или счётную подсистему, покрывающую почти всё множество E.

Здесь и далее длину отрезка или интервала D обозначаем $|D|$.

Доказательство. Будем считать, что E ограничено, скажем, $E \subset (0,1)$, и все отрезки покрытия лежат в интервале $(0,1)$ (общий случай получается рассмотрением каждого из множеств $E \cap (n, n+1)$, $n \in \mathbb{Z}$, и объединением соответствующих покрытий). Пусть уже выбраны непересекающиеся отрезки $d_1, \ldots, d_n$. Рассмотрим длины всех отрезков нашего покрытия, не пересекающихся с выбранными (на первом шаге — вообще всех). Пусть ℓ_n — точная верхняя грань этих длин. В качестве d_{n+1} возьмём отрезок покрытия, не пересекающийся с $d_1, \ldots, d_n$, длина которого больше $\ell_n/2$. Этот процесс либо оборвётся — и тогда покрыто уже всё множество E (это следует из того, что каждая точка в E покрыта сколь угодно малыми отрезками), либо даёт бесконечную последовательность отрезков $d_1, d_2, \ldots$ Докажем, что они покрывают почти всё множество E.

В самом деле, эти отрезки не пересекаются и лежат в $(0,1)$, так что сумма их длин не больше 1. Пусть D_n — интервал с той же серединой, что d_n, и в 5 раз большей длиной. Докажем, что при каждом k совокупность интервалов $D_k, D_{k+1}, \ldots$ покрывает множество $X = E \setminus \bigcup d_i$. Из этого будет следовать, что X имеет меру 0, поскольку ряд $\sum |d_n|$ сходится, а значит, $\sum_{n \geqslant k} |D_n| = 5 \sum_{n \geqslant k} |d_n| \to 0$.

Рассмотрим произвольную точку $x \in X$. Поскольку $x \notin \cup d_i$, найдётся отрезок покрытия $e \ni x$, не пересекающийся с $d_1, \ldots, d_{k-1}$. Пусть e не пересекается с $d_1, \ldots, d_m$ при некотором $m \geqslant k-1$. Тогда $2|d_{m+1}| > \ell_m \geqslant |e|$. Это неравенство показывает, что m при данном e не может быть сколь угодно велико, так что будем считать m максимально возможным. Иными словами, отрезок d_{m+1} уже пересекается с e, а тогда несложно видеть, что $e \subset D_{m+1}$ и, значит, $x \in D_{m+1}$. Это нам и нужно. □

Теперь докажем, что множество точек, в которых функция $f(x)$ не дифференцируема, имеет меру 0. Можно считать, что дело происходит не на всей оси, а на отрезке.

Наклоном функции f на отрезке $[u, v]$ назовём число $(f(v) - f(u))/(v - u)$.

Отсутствие производной возрастающей функции f в точке x_0 имеет одну из следующих причин.

1) Функция f имеет в точке x_0 бесконечную производную:

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = +\infty;$$

множество таких точек x_0 обозначим E_∞.

2) Найдутся такие рациональные числа $0 < p < q$, что для сколь угодно маленьких отрезков с одним из концов в точке x_0 наклон функции f бывает как меньше p, так и больше q. Множество соответствующих точек обозначим $E_{p,q}$.

Достаточно доказать, что меру 0 имеют множество E_∞ и каждое из множеств $E_{p,q}$.

В обоих случаях нам понадобится простая

Лемма. *Пусть на отрезке d выбрана система непересекающихся отрезков суммарной длины s, на каждом из которых наклон возрастающей функции f не меньше c. Тогда наклон функции f на отрезке d не меньше $c \cdot s/|d|$.*

Теперь рассмотрим множество E_∞. Выберем $M > 0$. Каждая точка множества E_∞ покрыта сколь угодно малым отрезком с наклоном функции f не меньше M. По теореме Витали почти всё множество E_∞ покрывается не более чем счётным объединением таких непересекающихся отрезков. Тогда в силу леммы сумма их длин стремится к 0 с ростом M, что и требовалось.

Для множества $E_{p,q}$ рассуждение аналогичное, но чуть тоньше. Именно, мы покажем, как из покрытия почти всего множества $E_{p,q}$ системой непересекающихся интервалов с некоторой суммой длин Q получить покрытие с суммой длин не больше $\frac{p}{q}Q$. Итерируя такой процесс, получим покрытие со сколь угодно малой суммой длин.

Итак, пусть почти всё множество $E_{p,q}$ покрыто интервалами $I_1, I_2 \ldots$ с суммой длин Q. Можно считать их непересекающимися. Внутри каждого интервала I_n применим теорему Витали к покрытию множества $E_{p,q} \cap I_n$ отрезками, наклон на которых меньше p. Затем рассмотрим интервалы, соответствующие этим отрезкам. Они не пересекаются и покрывают почти всё множество $E_{p,q}$. В каждом из них применим теорему Витали к отрезкам, наклон на которых не меньше q. Полученные интервалы снова покрывают почти всё $E_{p,q}$, и в силу леммы сумма их длин не больше, чем $\frac{p}{q}Q$, что и требовалось. (*Ф. Петров*)

12.9. Условие[1)]. В клетках бесконечной целочисленной решётки стоят целые числа. Докажите, что сумма чисел в некотором квадрате делится на 2008. (*С. В. Охитин, А. Я. Белов*)

Замечание. Упрощёнными вариантами данной задачи являются задача 3 старшего основного варианта весеннего тура XVII Турнира городов (1996 г.) и задачи А1 а, б) проекта «Почти-периодичность и символическая динамика» XVII Летней конференции Турнира городов (2005 г.).

[1)] Задача и её решение впервые опубликованы в [1].

Решение. Зафиксируем оси координат OX и OY, идущие по линиям сетки. Будем искать искомый квадрат в положительном квадранте. Каждой вершине решётки с координатами (x, y) мы сопоставим «цвет» $S(x, y)$, т. е. остаток по модулю 2008 суммы чисел, записанных в клетках с координатами (a, b), где $0 \leqslant a \leqslant x$, $0 \leqslant b \leqslant y$. Из двумерного обобщения теоремы ван дер Вардена [2, стр. 152] следует, что для любой конечной фигуры на решётке существует гомотетичная ей фигура одного цвета[2)]. В частности, существует квадрат с вершинами одного цвета. Пусть их координаты суть (x, y), $(x + k, y)$, $(x, y + k)$, $(x + k, y + k)$. Из совпадения величины S для вершин квадрата следует искомое, поскольку сумма чисел Σ в клетках квадрата, как легко убедиться, равна

$$\Sigma = S(x, y) - S(x + k, y) - S(x, y + k) + S(x + k, y + k) \equiv 0 \pmod{2008}.$$

Комментарий. Задача очевидным образом обобщается для k-мерного куба. Разбив куб на мелкие кубики, получаем аналогичный результат:

Теорема. *В клетках k-мерной решётки записаны целые числа. Тогда для любых m и n найдётся куб с ребром $L \cdot n$, где L — натуральное число, разбитый на n^k кубиков с ребром L, таких что сумма чисел в каждом из кубиков делится на m. (L, вообще говоря, зависит от k, n, m и расположения чисел.)*

Для одномерного случая это утверждение означает следующее: если на клетчатой ленте записаны целые числа, то найдётся n одинаковых отрезков, идущих подряд, на каждом из которых сумма чисел делится на m.

Другое обобщение заключается в следующем: в формулировке теоремы считаем, что записаны произвольные действительные числа, а в утверждении теоремы фразу *сумма чисел в каждом из кубиков делится на m* заменяем на фразу *отстоит от целого числа меньше, чем наперёд заданное число ε*. Доказательство повторяет те же соображения, только «цвета» получаются так: отрезок $[0, 1]$ разбиваем на $N > (n)^k \varepsilon^{-1}$ равных частей и «цветом» каждой точки с координатами $(x_1, \ldots, x_k)$ считаем номер части, куда попадает дробная часть суммы чисел, записанных в точках с координатами $(y_1, \ldots, y_k)$, где $1 \leqslant y_i \leqslant x_i$, $1 \leqslant i \leqslant k$.

Библиография

[1] *Белов А. Я., Охитин С. В.* Об одной комбинаторной задаче // УМН. 1993. Т. 48, вып. 2(290). С. 169–170.

[2)] Вопросы, связанные с теоремой ван дер Вардена, изложены в замечательной книжке Грэхема «Начала теории Рамсея» [3], которую мы рекомендуем читателю.

[2] *Бугаенко В. О.* Обобщённая теорема ван дер Вардена // Математическое просвещение. Сер. 3. Вып. 10. М.: МЦНМО, 2006. С. 151–160.
[3] *Грэхем Р.* Начала теории Рамсея. М.: Мир, 1984.

(*А. Я. Канель-Белов*)

13.6. Условие. Докажите, что в алгебре [квадратных] матриц порядка n выполняются следующие тождества.

а) Тождество Размыслова:

$$n\cdot\operatorname{tr}(A)\cdot\sum_{\sigma\in S_{n^2}}(-1)^\sigma x_{\sigma(1)}y_1x_{\sigma(2)}y_2x_{\sigma(3)}\cdot\ldots\cdot x_{\sigma(n^2-1)}y_{n^2-1}x_{\sigma(n^2)}=$$
$$=\sum_{\sigma\in S_{n^2}}(-1)^\sigma Ax_{\sigma(1)}y_1x_{\sigma(2)}y_2\cdot\ldots\cdot y_{n^2-1}x_{\sigma(n^2)}+$$
$$+\sum_{\sigma\in S_{n^2}}(-1)^\sigma x_{\sigma(1)}y_1Ax_{\sigma(2)}y_2\cdot\ldots\cdot y_{n^2-1}x_{\sigma(n^2)}+\ldots+$$
$$+\sum_{\sigma\in S_{n^2}}(-1)^\sigma x_{\sigma(1)}y_1x_{\sigma(2)}y_2\cdot\ldots\cdot y_{n^2-1}Ax_{\sigma(n^2)}.$$

б) Тождество Амицура — Левицкого:

$$\sum_{\sigma\in S_{2n}}(-1)^\sigma x_{\sigma(1)}x_{\sigma(2)}\cdot\ldots\cdot x_{\sigma(2n)}=0.$$

Здесь x_i, y_i, A — произвольные матрицы, S_k — группа перестановок k элементов, $(-1)^\sigma=+1$ для чётных перестановок и $(-1)^\sigma=-1$ для нечётных.

Решение. а) Напомним, что понятие *определителя* в учебниках линейной алгебры определяется тремя способами. Первый способ — через объём: с матрицей (a_{ij}) связывается линейное преобразование, переводящее вектор $\vec{e}_i$ в $\sum_i a_{ij}\vec{e}_j$, и $\det(a_{ij})$ есть коэффициент искажения объёмов. Второй способ определить это понятие — через известное равенство:

$$\det(a_{ij})=\sum_\sigma(-1)^\sigma\prod_i a_{i,\sigma(i)}.$$

Согласно третьему, аксиоматическому способу введения этого понятия, определитель есть полилинейная кососимметрическая функция на наборе векторов $\vec{v}_i$, равная 1 на наборе базисных векторов $\vec{e}_1$, $\vec{e}_2$, ...

Основное соображение, принадлежащее Ю. П. Размыслову, таково. Пусть $F(v_1,\ldots,v_m)$ — многочлен, полилинейный и кососимметричный относительно системы векторов v_i, а V — векторное пространство размерности m, порождённое векторами v_i. Тогда значение $F(v_1,\ldots,v_m)$ пропорционально

объёму параллелипипеда, образованного набором векторов $v_1, \ldots, v_m$, или, что то же самое, *определителю* матрицы, образованной векторами строк $v_1, \ldots, v_m$. Рассмотрим оператор $A \in \mathrm{End}(V)$. Пусть функция f полилинейна и кососимметрична относительно $\{x_i\}$ и произвольным образом зависит от y. Тогда для $x = (x_1, \ldots, x_m)$ имеем

$$f(A \cdot x, Y) = f(Ax_1, \ldots, Ax_m, Y) = \det(A) f(x, Y).$$

Теперь положим $A = E + ta$ и разложим по степеням t. В силу тождества Гамильтона — Кэли

$$f(A \cdot x, Y) = \left(\sum_{k=0}^{m} \Phi_k(a) t^k \right) f(x, Y), \tag{1}$$

где $\Phi_k(a)$ есть *форма k-го порядка* от оператора a. Она равна сумме главных миноров порядка k матрицы оператора[3)] a. В частности, $\Phi_1(a) = \mathrm{tr}(a)$, $\Phi_m(a) = \det(a)$, $\Phi_0(a) = 1$.

С другой стороны, коэффициент при t^k равен

$$\sum_{i_1 < \ldots < i_k} f(x_1, \ldots, x_m, Y)\big|_{ax_{i_\alpha} \to x_{i_\alpha}}$$

(здесь x_{i_α} заменяется на ax_{i_α}, а подстановка осуществляется в k местах, где появляется множитель t), в частности

$$\mathrm{tr}(a) f(x_1, \ldots, x_m, Y) = \sum_i f(x_1, \ldots, a \cdot x_i, \ldots, x_m, Y).$$

С матрицей A порядка n можно связать *оператор левого умножения* $L(A)$, переводящий каждую матрицу B в произведение AB. Покажем, что $\det(L(A)) = (\det(A))^n$. Пространство матриц есть прямая сумма своих подпространств R_i, отвечающих матрицам, у которых i-й столбец произвольный, а остальные — нулевые. Оператор $L(A)$ действует на каждом из этих подпространств стандартным образом, и все эти его ограничения имеют определитель, равный $\det(A)$. Поскольку на каждом из этих n подпространств объёмы изменяются в $\det(A)$ раз, итоговое изменение объёма окажется равным $(\det(A))^n$. Соответственно след $\mathrm{tr}(L(A))$ будет равен сумме следов на подпространствах, т. е. $n \cdot \mathrm{tr}(A)$.

Этот факт имеет альтернативное доказательство. Ясно, что $L(AB) = L(A)L(B)$, поэтому $\det(L(AB)) = \det(L(A)) \det(L(B))$. Если $A = \lambda \cdot E$,

[3)] То есть следу оператора $\bigwedge^k(a)$, действующего на векторном пространстве $\bigwedge^k(V)$. О внешнем произведении см. решение задачи 17.10, «Математическое просвещение», вып. 19, с. 263.

где E — единичная матрица, то $AB = \lambda B$, т. е. оператор умножения на A есть гомотетия в n^2-мерном пространстве $n \times n$-матриц. Его определитель равен $\lambda^{n^2} = (\lambda^n)^n = (\det(A))^n$. Если же $A(e_i) = e_i + e_j$, то A есть элементарная матрица с единичным определителем. С другой стороны, подстановка в кососимметрический многочлен $x_i = x_j$ приводит к нулевому результату, а из полилинейности тогда следует, что подстановка $x_i + x_j \to x_i$, $i \neq j$, результат не меняет, так что определитель $L(A)$ для такой элементарной матрицы A равен 1. Остаётся заметить, что произвольная матрица есть произведение элементарных.

В качестве f (см. начало решения) мы возьмём *многочлен Капелли*

$$C_n(x_1, \ldots, x_{n^2}, y_1, \ldots, y_{n^2-1}) = \sum_{\sigma \in S_{n^2}} (-1)^\sigma x_{\sigma(1)} y_1 \ldots y_{n^2-1} x_{\sigma(n^2)};$$

в качестве V — пространство всех $n \times n$-матриц, так что $m = n^2$. Тогда

$$\begin{aligned} C_n(A \cdot x_1, \ldots, A \cdot x_{n^2}, y_1, \ldots, y_{n^2-1}) = \\ = \det(L(A)) C_n(x_1, \ldots, x_{n^2}, y_1, \ldots, y_{n^2-1}) = \\ = \det(A)^n C_n(x_1, \ldots, x_{n^2}, y_1, \ldots, y_{n^2-1}). \end{aligned}$$

И соответственно

$$\sum_i C_n(x_1, \ldots, ax_i, \ldots, x_{n^2}, y_1, \ldots, y_{n^2-1}) = n \operatorname{tr}(a) C_n(x_1, \ldots, x_{n^2}, y_1, \ldots, y_{n^2-1}),$$

что нам и требовалось установить.

Замечание. Рассмотрим алгебру, в которой выполняется тождество Капелли порядка $k+1$, т. е. все многочлены, полилинейные и кососимметрические по $k+1$ переменным, обращаются в ноль (в матричном случае $k = n^2$). Пусть $C_n(x_1, \ldots, x_k, \vec{y})$ — многочлен Капелли порядка k. Определим операторы

$$\delta_m(z)(f) = \sum_{i_1 < \ldots < i_m} C_n(x_1, \ldots, x_k, \vec{y})\big|_{x_{i_\alpha} z \to x_{i_\alpha}}.$$

При этом естественно ожидать, что операторы $\delta_m(z)$ будут вести себя аналогично матричному случаю. И это действительно так. Если многочлен $F(x_1, \ldots, x_n, Y, t)$ полилинеен относительно набора $\{x_1, \ldots, x_n, t\}$ и косоциметричен относительно $\{x_1, \ldots, x_n\}$ (и произвольным образом зависит от Y), то имеет место равенство

$$F(x_1, \ldots, x_n, Y, z^n) = \sum (-1)^m \delta_m(z)(f).$$

Это равносильно умножению на следы элементов, а добавление следов ведёт к конечномерности над кольцом, порождённым следами. Элементы алгебры оказываются алгебраическими над операторами $\delta_m(f)$ и тем самым конечномерными над ними. Это соображение оказалось необходимым в теории тождеств, в частности, для решения проблемы Шпехта.

б) Пусть $F(x_1, \dots, x_n)$ — функция, линейная по каждому x_i. *Оператором альтернирования* Alt назовём оператор

$$\mathrm{Alt}(F)\colon F(x_1, \dots, x_n) \to \sum_{\sigma \in S_n} (-1)^\sigma F(x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Ясно, что

$$\mathrm{Alt}(F + G) = \mathrm{Alt}(F) + \mathrm{Alt}(G). \tag{2}$$

Кроме того, если F, G — полилинейные многочлены от непересекающихся множеств из k и $n-k$ переменных и $\mathrm{Alt}(F) = 0$, то

$$\mathrm{Alt}(F \cdot G) = 0. \tag{3}$$

В самом деле, пусть H — подгруппа перестановок из S_n, не смешивающих переменные x_i и y_i. Для каждого разбиения γ набора из n переменных $\{x_i, y_j\}$ на две группы из k и $n-k$ переменных (обозначим их Z^γ, T^γ) зафиксируем нумерацию переменных в каждой группе и соответственно элемент $\tau_\gamma\colon x_i \to z_i^\gamma, y_j \to t_j^\gamma$. Легко видеть, что тогда

$$\begin{aligned}\mathrm{Alt}(F \cdot G) &= \sum_\gamma (-1)^{\tau_\gamma} \sum_{\sigma_1 \in S_k} (-1)^{\sigma_1} F\big(z^\gamma_{\sigma_1(1)}, \dots, z^\gamma_{\sigma_1(k)}\big) \times \\ &\quad \times \sum_{\sigma_2 \in S_{n-k}} (-1)^{\sigma_2} G\big(t^\gamma_{\sigma_2(1)}, \dots, t^\gamma_{\sigma_2(n-k)}\big) = \\ &= \sum_\gamma (-1)^{\tau_\gamma} \cdot 0 \cdot \sum_{\sigma_2 \in S_{n-k}} (-1)^{\sigma_2} G\big(t^\gamma_{\sigma_2(1)}, \dots, t^\gamma_{\sigma_2(n-k)}\big) = 0.\end{aligned}$$

Рассмотрим моном $M = x_1 \dots x_{2k}$. Каждой чётной перестановке σ и члену $x_{\sigma(1)} \dots x_{\sigma(2k-1)} x_{\sigma(2k)}$ сопоставляется член $x_{\sigma(2k)} x_{\sigma(1)} \dots x_{\sigma(2k-1)}$ и соответствующая нечётная перестановка. При этом $\mathrm{Alt}(M)$ есть линейная комбинация выражений вида

$$x_{\sigma(1)} \dots x_{\sigma(2k-1)} x_{\sigma(2k)} - x_{\sigma(2k)} x_{\sigma(1)} \dots x_{\sigma(2k-1)} = [x_{\sigma(2k)}, x_{\sigma(1)} \dots x_{\sigma(2k-1)}],$$

где $[a, b] = ab - ba$. Поскольку $\mathrm{tr}(AB) = \mathrm{tr}(BA)$, имеем

$$\mathrm{tr}(\mathrm{Alt}(x_1 \dots x_{2k})) = 0. \tag{4}$$

Аналогично *оператор симметрирования* Sym действует как

$$\mathrm{Sym}(F)\colon F(x_1,\dots,x_n)\to\sum_{\sigma\in S_n}F(x_{\sigma(1)},\dots,x_{\sigma(n)}).$$

Он обладает свойствами, аналогичными (2) и (3).

Пусть $F(x,Y)$ — многочлен, однородный степени k относительно переменной x и произвольным образом зависящий от группы переменных Y, причём $Y\not\ni x$. Подставим $\sum_{i=1}^k x_i$ вместо x (при этом переменные x_i не входят в Y) и рассмотрим члены, линейные по каждому x_i, — они входят по разу, пусть их сумма равна G. Она называется (*полной*) *линеаризацией* многочлена F.

Лемма. *Функция G симметрична относительно $\{x_i\}$. Если $F(x,Y)=H(x_1,\dots,x_k,Y)|_{x\to x_i}$, где H полилинейно по $x_1,\dots,x_k$, то $G=\mathrm{Sym}(H)$.*

Пусть теперь A — произвольная матрица порядка n. Тогда по теореме Гамильтона — Кэли A удовлетворяет характеристическому уравнению

$$A^n-\xi_1(A)+\ldots+(-1)^k\xi_k(A)+\ldots+(-1)^n\xi_n(A)=0. \tag{5}$$

Здесь $\xi_1(A)=\mathrm{tr}(A)$, $\xi_n(A)=\det(A)$, и если A диагонализуема и имеет собственные числа $\lambda_1,\dots,\lambda_n$, то k-й коэффициент характеристического многочлена $\xi_k(A)$ равен $\sum_{i_1<\ldots<i_k}\lambda_{i_1}\dots\lambda_{i_k}$. Положим $s_m=\sum_{i=1}^n\lambda_i^m$.

По основной теореме о симметрических многочленах, $\xi_k=P_k(s_1,\dots,s_n)$ для некоторых многочленов P_k, так что $\xi_k(A)=P_k(\mathrm{tr}(A),\dots,\mathrm{tr}(A^k))$. Например, $P_2(x,y)=(x^2-y^2)/2$ и $\xi_2(A)=(\mathrm{tr}^2(A)-\mathrm{tr}(A^2))/2$. Перепишем уравнение (5) в виде

$$A^n=-\sum_{k=1}^n(-1)^kP_k(\mathrm{tr}(A),\dots,\mathrm{tr}(A^k))A^{n-k}. \tag{6}$$

Многочлены от матриц и следов их произведений называются *многочленами со следом*. В нашем случае возникают мономы со следом вида $M=A^{n-k}\,\mathrm{tr}(A^{i_1})\dots\mathrm{tr}(A^{i_s})$, где $\sum i_\alpha=k$, линейные комбинации которых и порождают все $P_k(\mathrm{tr}(A),\dots,\mathrm{tr}(A^k))A^{n-k}$.

Подставим вместо A сумму $\sum_{i=1}^n A_i$ и рассмотрим возникающие члены, линейные относительно каждого из A_i. Согласно лемме моному A^n отвечает сумма $\sum_{\sigma\in S_n}A_{\sigma(1)}\dots A_{\sigma(n)}$, а моному $M=A^{n-k}\,\mathrm{tr}(A^{i_1})\dots\mathrm{tr}(A^{i_s})$ будет отвечать сумма

$$S_M=\sum_{\sigma\in S_n}A_{\sigma(1)}\dots A_{\sigma(n-k)}\times$$
$$\times\,\mathrm{tr}(A_{\sigma(n-k+1)}\dots A_{\sigma(n-k+i_1)})\dots\mathrm{tr}(A_{\sigma(n-k+i_1+\ldots+i_{s-1})}\dots A_{\sigma(n-k+i_1+\ldots+i_s)}).$$

Теперь положим $A_k = x_{2k-1}x_{2k}$, получим выражение $\widehat{S}_M$ и проальтернируем результат по x_j. Таким образом, сумма $\sum_{\sigma\in S_n} A_{\sigma(1)}\dots A_{\sigma(n)}$ превратится в $n!\cdot \mathrm{Alt}_{2n}(x_1,\dots,x_{2n})$.

Из формул (2)–(4) следует, что каждому моному M, содержащему следы, будет отвечать ноль.

Но тогда $n!\cdot \mathrm{Alt}_{2n}(x_1,\dots,x_{2n})$, в силу равенства (6), оказывается линейной комбинацией нулей, поэтому $\mathrm{Alt}_{2n}(x_1,\dots,x_{2n}) = 0$ будет тождеством, что и требуется. Теорема Амицура — Левицкого доказана.

Замечание. Линеаризация многочленов $\xi_k(A)$ устроена следующим образом. Рассмотрим перестановку σ переменных x_i, она состоит из циклов. Каждому такому циклу $a_{i_1}\to a_{i_2}\to\dots\to a_{i_s}\to a_{i_1}$ сопоставим следовый моном $\mathrm{tr}(a_{i_1}\dots a_{i_k})$, а перестановке σ — их произведение, обозначим его A_σ. Результат линеаризации $\xi_k(A)$ есть $\sum_{\sigma\in S_k}(-1)^\sigma A_\sigma$. Если тождество Гамильтона — Кэли для матриц порядка n умножить на x_0, взять от него след и линеаризовать, то в матрицах порядка $n+1$ этому соответствует линеаризация многочлена ξ_n. (*А. Я. Канель-Белов*)

15.5. Условие. Даны две бесконечно дифференцируемые функции

$$f(x) = x + a_2x^2 + a_3x^3 + \dots,$$
$$g(x) = x + b_2x^2 + b_3x^3 + \dots$$

Известно, что $f(g(x)) = x$ и что все числа $k!a_k$ — целые. Докажите, что все числа $k!b_k$ — тоже целые. (*Е. Ю. Буньковa*)

Решение. Из условия следует, что значение k-й производной функции $f(x)$ в нуле целое ($k = 0, 1, 2, \dots$), и нужно доказать, что значение k-й производной функции $g(x)$ в нуле также целое ($k = 0, 1, 2, \dots, n$). Воспользуемся индукцией по k. База очевидна. Значение k-й производной функции $f(g(x))$ равно $f'(g(x))g^{(k)}(x) + P(x)$, где в $P(x)$ входят разнообразные слагаемые из произведений $f^{(l)}(g(x))$, $1\leqslant l\leqslant k$, и степеней m-х производных функции $g(x)$ при $1\leqslant m < k$. При $x = 0$ это выражение с учётом равенств $g(0) = 0$ и $f'(x)|_{x=0} = 1$ принимает вид $g^{(k)}(0) + P(0)$, причём $P(0)$ целое по предположению индукции. С другой стороны, ввиду условия $f(g(x)) = x$ это выражение равно 1 при $k = 1$ и 0 при $k > 1$, т. е. также целое. Отсюда и следует доказываемое утверждение. (*Д. Максимов, Ф. Петров*)

16.3. Условие. Пусть у функции, определённой на отрезке [или на прямой], в каждой точке этого отрезка [прямой] есть конечный предел (не обязательно совпадающий со значением в точке). Насколько такая функция может отличаться от непрерывной? Более точно: каким может быть множество точек разрыва у такой функции? (*М. Прасолов*)

Ответ: множеством точек разрыва такой функции может служить любое не более чем счётное множество точек отрезка [прямой].

Решение. Пусть у функции f на отрезке I в каждой точке есть предел. Положим функцию g в точке x отрезка I равной пределу функции f в точке x. Докажем, что функция g непрерывна и что множество точек, где f отличается от g больше, чем на произвольное фиксированное число $c > 0$, конечно.

Пусть $\varepsilon > 0$. Для точки x_0 на отрезке I выберем такую окрестность U, в которой f отличается от $g(x_0)$ не больше, чем на ε, за исключением, быть может, самой точки x_0. Тогда в U функция g тоже отличается от $g(x_0)$ не больше, чем на ε, потому что в каждой точке окрестности U функция g равна пределу функции f. Непрерывность g доказана.

Предположим, что множество точек x, где $|f(x) - g(x)| > c$, бесконечно. Тогда оно имеет предельную точку x_0. Выберем в этом множестве последовательность $x_n \to x_0$. Тогда

$$0 = g(x_0) - g(x_0) = \lim_{x_n \to x_0} f(x_n) - \lim_{x_n \to x_0} g(x_n) = \lim_{x_n \to x_0} (f(x_n) - g(x_n)).$$

Но последний предел не может равняться нулю, потому что последовательность состоит из чисел, больших чем $c > 0$. Значит, указанное множество конечно.

Поскольку любая точка разрыва функции f принадлежит такому множеству для некоторого c, множество этих точек не более чем счётно.

Теперь возьмём непрерывную функцию g на отрезке I, последовательность a_n, предел которой равен 0, счётное подмножество x_n отрезка I и определим функцию

$$f(x) = \begin{cases} g(x), & x \neq x_n, \\ g(x_n) + a_n, & x = x_n. \end{cases}$$

Тогда f имеет в каждой точке предел, причём любая функция, имеющая в каждой точке предел, может быть получена таким способом, что следует из доказательства, приведённого выше. В качестве примера такой функции подойдёт функция Римана, равная 0, если значение аргумента иррационально, и q, если оно представимо несократимой дробью со знаменателем q.

Полученные результаты справедливы и для функции, заданной на прямой, поскольку прямая есть объединение счётного множества отрезков.

(*М. Прасолов*)

17.1. Условие. Можно ли в куб достаточно большой размерности с ребром 1 см вложить здание МГУ? (*Ф. Ивлев*)

Ответ: да, можно.

Решение. Рассмотрим, например, размерность $3\,000\,000\,000\,000 = 3 \cdot 10^{12}$ и положим наш куб одной вершиной в начало координат, а рёбрами параллельно осям координат. Рассмотрим четыре вершины: точку A с нулевыми координатами, точку B, первые $2 \cdot 10^{12}$ координат которой равны 1, а остальные равны 0, точку C, у которой последние два триллиона координат равны 1, а остальные нулю, и, наконец, точку D, у которой первый триллион координат равен единице, следующий триллион координат равен нулю, а затем опять идёт триллион единиц. Попарные расстояния между любыми двумя из этих точек по многомерной теореме Пифагора равны $\sqrt{2 \cdot 10^{12}}$. Следовательно, эти четыре точки образуют правильный тетраэдр с ребром, равным $\sqrt{2} \cdot 10^6$. А это больше, чем 14 км. Очевидно, что Главное здание МГУ можно поместить в правильный тетраэдр с таким большим ребром, а весь этот тетраэдр лежит внутри нашего куба.

Можно было рассматривать не тетраэдр, а куб, порождённый векторами $(a; 0; 0)$, $(0; a; 0)$ и $(0; 0; a)$ (здесь вместо каждого нуля надо подставить 10^{12} нулей, а вместо $a - 10^{12}$ единиц). Тогда он тоже весь лежит внутри нашего гиперкуба, при этом его ребро по длине равно 10 км. (*Ф. Ивлев*)

19.3. Условие. Решите уравнения

$$\text{а) } x^2 + y^2 = (x+1)^3; \quad \text{б) } x^2 + xy + 2y^2 = (y+1)^3$$

в целых числах. (*Н. Н. Осипов*)

Ответ:
а) $(x, y) \in \{(0, \pm 1), (88, \pm 835)\}$;
б) $(x, y) \in \{(-39, 10), (-3, 1), (-1, 0), (1, 0), (2, 1), (29, 10)\}$.

Решение. а) При $x < 0$ решений нет, поэтому далее считаем $x \geqslant 0$.

Числа x и y взаимно просты, причём y нечётно, а x чётно (если x нечётно, то $x^2 + y^2 \equiv 2 \pmod 4$, а $(x+1)^3 \equiv 0 \pmod 4$). Запишем уравнение в виде

$$(x + y\sqrt{-1})(x - y\sqrt{-1}) = (x+1)^3 \qquad (*)$$

и воспользуемся тем, что кольцо

$$\mathbb{Z}[\sqrt{-1}] = \{a + b\sqrt{-1} \colon a, b \in \mathbb{Z}\}$$

целых гауссовых чисел *факториально* (доказательство, сводящееся к проверке *евклидовости*, можно найти, например, в учебнике [1, стр. 114–115]).

Числа $x \pm y\sqrt{-1}$ взаимно просты в кольце $\mathbb{Z}[\sqrt{-1}]$ тогда и только тогда, когда целые числа x и y взаимно просты и имеют разную чётность (упражнение для читателя). Из факториальности $\mathbb{Z}[\sqrt{-1}]$ и равенства $(*)$

следует, что

$$x + y\sqrt{-1} = \varepsilon(a + b\sqrt{-1})^3$$

для некоторых $a,\ b \in \mathbb{Z}$ и *единицы* (обратимого элемента) $\varepsilon \in \mathbb{Z}[\sqrt{-1}]$. Но все единицы $\mathbb{Z}[\sqrt{-1}]$ суть $\pm 1, \pm\sqrt{-1}$, и каждая из них является кубом единицы, что позволяет нам считать $\varepsilon = 1$. Таким образом, имеем

$$x = a^3 - 3ab^2, \quad y = 3a^2b - b^3.$$

Кроме того,

$$(x+1)^3 = x^2 + y^2 = (a + b\sqrt{-1})^3(a - b\sqrt{-1})^3 = (a^2 + b^2)^3,$$

и, так как $x+1$ и a^2+b^2 вещественны, $x+1 = a^2+b^2$. Исключив x, получим уравнение

$$a^3 - 3ab^2 - a^2 - b^2 + 1 = 0. \qquad (**)$$

Оно легко решается, поскольку может быть переписано в виде

$$27b^2 = 9a^2 - 12a + 4 + \frac{23}{3a+1}.$$

Отсюда $(a, b) \in \{(-8, \pm 5), (0, \pm 1)\}$ и $(x, y) \in \{(0, \pm 1), (88, \pm 835)\}$.

б) Рассуждаем так же, как и в п. а), только используем факториальность кольца

$$\mathbb{Z}[\omega] = \{a + b\omega \colon a, b \in \mathbb{Z}\}, \quad \omega = \frac{1 + \sqrt{-7}}{2}$$

(это кольцо является также и евклидовым, предпоследним в списке евклидовых колец целых чисел мнимых квадратичных полей). Получим

$$x = a^3 - 6ab^2 - 2b^3, \quad y = 3a^2b - b^3 + 3ab^2, \quad y + 1 = a^2 + ab + 2b^2$$

для некоторых $a, b \in \mathbb{Z}$, откуда

$$3a^2b - b^3 + 3ab^2 - a^2 - ab - 2b^2 + 1 = 0.$$

Положив $u = 2a + b$, получим

$$3bu^2 - 7b^3 - u^2 - 7b^2 + 4 = 0. \qquad (***)$$

Это уравнение решается так же, как и уравнение $(**)$.

Комментарий. Уравнения $(**)$ и $(***)$ относятся к специальному типу уравнений, которые могут быть решены *методом Рунге* (см., например, [3, стр. 11–13]). В статье [2] предложена элементарная версия этого метода для кубических уравнений. В общем случае решение кубических уравнений методом Рунге будет сложнее из-за наличия линейных слагаемых (в уравнениях $(**)$ и $(***)$ они отсутствуют).

Библиография

[1] *Винберг Э. Б.* Курс алгебры. 2-е изд. М.: МЦНМО, 2013.
[2] *Осипов Н. Н.* Элементарная версия метода Рунге для кубических уравнений // Математика в школе. 2012. №1. С. 64–69.
[3] *Спринджук В. Г.* Классические диофантовы уравнения от двух неизвестных. М.: Наука, 1982.

(*Н. Н. Осипов*)

19.7. Условие. Докажите неравенство $\dfrac{R}{r} < \dfrac{\pi}{\alpha\beta\gamma}$, где R, r — радиусы описанной и вписанной окружностей треугольника, α, β, γ — его углы в радианах. (*К. Э. Каибханов*)

Решение. Из формулы

$$r = \frac{S}{p} = \frac{abc}{4pR},$$

теоремы синусов и формулы суммы синусов следует, что

$$r = 4R \sin\frac{\alpha}{2} \sin\frac{\beta}{2} \sin\frac{\gamma}{2}.$$

Поэтому искомое неравенство можно переписать в виде

$$\frac{\sin x}{x} \cdot \frac{\sin y}{y} \cdot \frac{\sin z}{z} > \frac{2}{\pi},$$

где $x = \alpha/2$, $y = \beta/2$, $z = \gamma/2$.

Но функция $f(t) = \ln \dfrac{\sin t}{t}$ на интервале $(0, \pi/2)$ вогнута. Значит, минимум $f(x) + f(y) + f(z)$ в треугольнике $\{x, y, z \geqslant 0, x + y + z = \pi/2\}$ достигается в его вершинах, откуда и получаем нужное неравенство.

(*А. Заславский*)

19.9. Условие. Дан треугольник ABC и такие положительные числа p, q, что

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Пусть X — точка плоскости, для которой сумма $(AX)^p + (BX)^p + (CX)^p$ минимальна, а A', B', C' — точки на сторонах BC, CA, AB, для которых сумма $(A'X)^q + (B'X)^q + (C'X)^q$ минимальна. Докажите, что $CX \perp A'B'$. (*А. Заславский*)

Решение. Подробное решение этой задачи содержится в статье [2]. Приведём его краткое изложение.

Прежде всего докажем, что точки A', B', C' являются проекциями некоторой точки на стороны треугольника. Действительно, окружности $AB'C'$, $BC'A'$ и $CA'B'$ пересекаются в некоторой точке Y, поэтому $A'B' \geqslant CY \sin C$,

$B'C' \geqslant AY \sin A$, $C'A' \geqslant BY \sin B$, причём равенства достигаются тогда и только тогда, когда A', B', C' — проекции Y на стороны. Следовательно, задача нахождения точек A', B', C' сводится к определению точки Y, минимизирующей функцию $g(Y) = (a|AY|)^q + (b|BY|)^q + (c|CY|)^q$, где $a = |BC|$, $b = |AC|$, $c = |AB|$. Заметим теперь, что если O — фиксированная точка плоскости, X — переменная точка, а $p > 1$, то функция $|OX|^p$ выпуклая, а её градиент равен $p|OX|^{p-2}\overrightarrow{OX}$. Отсюда следует, что градиент функций $f(X) = |XA|^p + |XB|^p + |XC|^p$ и $g(Y)$ на границе треугольника направлен во внешнюю сторону, т. е. минимумы этих функций достигаются во внутренних точках и, соответственно, их градиенты в этих точках обращаются в нуль. Отсюда получаем, что барицентрические координаты точки X равны $(|XA|^{p-2}, |XB|^{p-2}, |XC|^{p-2})$, а барицентрические координаты точки Y — $(a^q|YA|^{q-2}, b^q|YB|^{q-2}, c^q|YC|^{q-2})$. При $1/p + 1/q = 1$ это означает, что X и Y изогонально сопряжены, что равносильно утверждению задачи.

Примечание. Подробнее о рассмотренных экстремальных задачах можно прочесть в статье [3], а об изогональном сопряжении — в статье [1].

Библиография

[1] *Акопян А., Заславский А.* Разные взгляды на изогональное сопряжение // Математическое просвещение. Сер. 3. Вып. 11. М.: МЦНМО, 2007. С. 61–78.

[2] *Заславский А.* Ещё раз о пространствах L_p и замечательных точках треугольника // Квант. 2013. № 5–6. С. 45.

[3] *Протасов В., Тихомиров В.* О пространствах L_p и замечательных точках треугольника // Квант. 2012. № 2. С. 2–11.

(*А. Заславский*)

Указатель условий, решений и статей по мотивам задач из «Математического просвещения»

(Если в графе «Условие» номер выпуска больше, чем в графе «Задача», то он относится к исправленному условию.)

Задача	Условие	Решение или статья по мотивам	Задача	Условие	Решение или статья по мотивам
1.1	1, 193	6, 137–139	3.8	3, 233	6, 148–149
1.2	1, 193	5, 218	3.9	8, 247	15, 206–211
1.3	1, 193	4, 218	3.10	3, 233	8, 239–245
1.4	1, 193	5, 218–221	3.11	3, 233	8, 186–221
1.5	1, 194	6, 139–140	4.1	4, 215	6, 149–150
1.6	1, 194	5, 221–223; 10, 274	4.2	4, 215	7, 193
1.7	1, 194	4, 219	4.3	4, 215	7, 193–194
1.8	1, 194	4, 220	4.4	4, 216	7, 194
1.9	1, 194	5, 223–225	4.5	4, 216	6, 150
1.10	1, 194	4, 220; 5, 225–227	4.6	4, 216	6, 150–151
2.1	2, 216	4, 221	4.7	4, 216	6, 151–152
2.2	2, 216	5, 227–228	4.8	4, 216	5, 228–229
2.3	2, 216	4, 221	4.9	4, 217	8, 237–238
2.4	2, 217	4, 221–222	4.10	4, 217	7, 194–195
2.5	2, 217	4, 222	4.11	4, 217	8, 249–252; 15, 212–218
2.6	2, 217	4, 222–223	4.12	4, 217	18, 263–269
2.7	2, 217	6, 140–142	5.1	5, 216	6, 152–153
2.8	2, 217	6, 142–145; 18, 259–261	5.2	6, 135	7, 195–196
2.9	2, 217	6, 145–147	5.3	5, 216	6, 153
2.10	6, 135	8, 248–249	5.4	5, 216–217	7, 196–198
3.1	3, 232	4, 223	5.5	5, 217	8, 252–254
3.2	3, 232	4, 223–224	5.6	5, 217	14, 275–276
3.3	3, 232	5, 227–228; 19, 241–247	5.7	5, 217	6, 153
3.4	3, 232	4, 224	5.8	5, 217	7, 198
3.5	3, 233	4, 225; 18, 262	5.9	5, 217	10, 232–235, 236–242; 11, 145–148
3.6	3, 233	7, 190–193	5.10	5, 217	8, 254
3.7	3, 233	6, 147–148	6.1	6, 133	8, 255

Задача	Условие	Решение или статья по мотивам	Задача	Условие	Решение или статья по мотивам
6.2	6, 133	9, 225	9.10	9, 224	10, 265–272
6.3	6, 133	9, 225–226	9.11	9, 224	
6.4	6, 133	9, 226–227	10.1	10, 278	13, 185–186
6.5	6, 133–134		10.2	10, 278	11, 172
6.6	6, 134	8, 255–256	10.3	10, 278	
6.7	6, 134	11, 165–166	10.4	10, 278–279	11, 173–174
6.8	6, 134	8, 256–257	10.5	10, 279	14, 240–255
6.9	6, 134	8, 222–228; 229–236	10.6	10, 279	12, 238–239
6.10	6, 134	20, 252–253	10.7	10, 279	
6.11	6, 134	9, 227–229; 17, 198–199	10.8	10, 279	
6.12	6, 134	8, 258–259	10.9	10, 279	14, 277–278
7.1	7, 187	9, 229	10.10	10, 280	11, 149–158
7.2	7, 187	8, 259–260	10.11	10, 280	
7.3	7, 187	9, 229; 10, 274; 10, 275	10.12	10, 280	
7.4	7, 187	9, 229–230	11.1	11, 162	13, 186; 14, 278
7.5	7, 187–188		11.2	11, 162	13, 186–189
7.6	7, 188	16, 233–235	11.3	11, 162	17, 199
7.7	7, 188		11.4	11, 162–163	
7.8	7, 188	13, 182–184	11.5	11, 163	
7.9	7, 189	9, 230–232	11.6	11, 163	14, 279
7.10	11, 164		11.7	11, 163	
7.11	7, 189	9, 232–233	11.8	11, 163	
7.12	7, 189	11, 166–168	11.9	11, 163	
8.1	8, 246	13, 184	11.10	11, 164	17, 199–200
8.2	8, 246	10, 281	11.11	11, 164	
8.3	8, 246	10, 281–282	11.12	11, 164	
8.4	8, 246	10, 243–264, 282–284	12.1	12, 235	15, 236–237; 17, 187–188
8.5	8, 246	9, 215–217	12.2	12, 235	
8.6	8, 246		12.3	12, 235	16, 236; 17, 200
8.7	8, 247		12.4	12, 235	14, 279–280
8.8	8, 247	9, 233	12.5	12, 236	
8.9	8, 247	11, 168–169	12.6	12, 236	20, 254–256
8.10	8, 247		12.7	12, 236	16, 236–237
8.11	8, 247		12.8	12, 236	14, 256–269
8.12	8, 247	20, 238–242	12.9	12, 236	20, 256–258
9.1	9, 223	14, 276–277	12.10	12, 236	13, 189–190
9.2	9, 223	13, 184–185	12.11	12, 236	
9.3	9, 223		12.12	12, 236	
9.4	9, 223	10, 284–285	13.1	13, 179	14, 280–281
9.5	9, 223	14, 277	13.2	13, 179	14, 270–271; 14, 281
9.6	9, 223–224	12, 237–238	13.3	13, 179	15, 237
9.7	9, 224		13.4	13, 179	
9.8	9, 224	12, 238	13.5	13, 179	
9.9	9, 224	11, 169–172	13.6	13, 179–180	20, 258–263

Задача	Условие	Решение или статья по мотивам	Задача	Условие	Решение или статья по мотивам
13.7	13, 180		16.10	16, 231	
13.8	13, 180		16.11	16, 231	17, 203–207
13.9	13, 180	19, 249–253	16.12	16, 232	
13.10	13, 181	20, 215–227	17.1	17, 196	20, 264–265
13.11	13, 181		17.2	17, 196	
13.12	13, 181		17.3	17, 196	
14.1	14, 272		17.4	17, 196	
14.2	14, 272		17.5	17, 196	
14.3	14, 272	15, 237–239	17.6	17, 196	
14.4	14, 272		17.7	17, 197	
14.5	14, 272	15, 239–240	17.8	17, 197	
14.6	14, 273	15, 240–241	17.9	17, 197	
14.7	15, 235	15, 241–244	17.10	17, 197	19, 263
14.8	15, 235	15, 219–228	17.11	17, 197	20, 228–237
14.9	14, 273	16, 237–239	17.12	17, 197	
14.10	14, 273		18.1	18, 255	
14.11	14, 274		18.2	18, 255–256	19, 264
14.12	14, 274	15, 229–230	18.3	18, 256	
15.1	15, 232		18.4	18, 256	
15.2	15, 232		18.5	18, 257	
15.3	15, 232		18.6	18, 257	
15.4	15, 232		18.7	18, 257	19, 264–266
15.5	15, 233	20, 263	18.8	18, 257	19, 254–256
15.6	15, 233		18.9	18, 257	
15.7	15, 233		18.10	18, 258	19, 266–267
15.8	15, 233		18.11	18, 258	
15.9	15, 233		18.12	18, 258	
15.10	15, 234	17, 200–202	19.1	19, 257	
15.11	15, 234		19.2	19, 257	
15.12	15, 234		19.3	19, 257	20, 265–267
16.1	16, 230	18, 269	19.4	19, 258	
16.2	16, 230	17, 202–203	19.5	19, 258	
16.3	16, 230	20, 263–264	19.6	19, 258	
16.4	16, 230	19, 260	19.7	19, 258	20, 267
16.5	16, 230		19.8	19, 258	
16.6	16, 231	17, 192–195	19.9	19, 258	20, 267–268
16.7	16, 231		19.10	19, 258	
16.8	16, 231		19.11	19, 258	
16.9	16, 231	19, 260–262	19.12	19, 259	

Информация для авторов

Сборник «Математическое просвещение» предназначен для широкого круга математиков — исследователей и преподавателей, а также студентов, старшеклассников и всех, кто интересуется математикой. Издание публикует материалы по различным областям математики, а также по вопросам её истории и преподавания, интересные и доступные указанной аудитории. Сборник «Математическое просвещение» не имеет возможности публиковать статьи, содержащие существенно новые научные результаты, оценка которых доступна лишь специалистам в данной области. Он также не публикует материалы по текущим вопросам преподавания математики в школе и вузах.

Сборник выходит, как правило, один раз в год, в марте–апреле. Редакция просит авторов представлять материалы по электронной почте на адрес matpros@mccme.ru в виде файлов pdf и tex (с файлами рисунков и т. п., если требуется). Допускается присылка статей, набранных в Word.

Просим авторов указывать в конце статьи: свою фамилию с инициалами, место работы/обучения, электронный адрес для дальнейшей переписки.

Редакция будет благодарна авторам, если они оформят ссылки на литературу аналогично тому, как это сделано в опубликованных выпусках сборника http://www.mccme.ru/free-books/matpros.html

Авторы опубликованных статей имеют право на 2 экземпляра статьи каждый. По поводу их получения обращайтесь на matpros@mccme.ru

Издательство Московского центра
непрерывного математического образования

119002, Москва, Большой Власьевский пер., 11. Тел. (499) 241–08–04.

Отпечатано в ППП «Типография „Наука“».
121099, Москва, Шубинский пер., 6.

Подписано в печать 14.03.2016 г. Формат $70\times100\,{}^{1}\!/_{16}$. Бумага офсетная.
Печать офсетная. Печ. л. 17. Тираж 1000 экз. Заказ №

Книги издательства МЦНМО можно приобрести в магазине «Математическая книга»,
Москва, Большой Власьевский пер., 11. Тел. (495) 745–80–31.
E-mail: biblio@mccme.ru, http://biblio.mccme.ru