МАТЕМАТИЧЕСКОЕ ПРОСВЕЩЕНИЕ

Третья серия

выпуск 2

МЦНМО 1998

Адрес редакции:
121002, Москва, Б. Власьевский пер., д.11, к. 211
(с пометкой «Математическое просвещение»)

Email: matpros@mccme.ru

В этом — втором — сборнике новой серии «Математического просвещения» в разделе, посвящённом проблемам современной математики, помещён цикл статей о математических вопросах криптографии; раздел «Математический мир» составлен из статей, посвящённых деятельности Н. Бурбаки, первым филдсовским медалям и выдающимся результатам А. Н. Колмогорова и Л. С. Понтрягина тридцатых годов, которые не рассматривались филдсовским комитетом по нематематическим причинам. В разделе «Проблемы математического образования» представлена программа «Матшкольник», содержащая основные сведения из математики, которые должны знать выпускники математических школ; причём уровень требований по каждой теме задаётся образцами задач. Среди материалов по истории — размышления о первых московских математических олимпиадах. Помимо этого, ряд статей посвящены ярким и интересным математическим фактам и миниатюрам — различным доказательствам выпуклости чебышёвских множеств, полному решению обобщённой задачи Мальфатти, неожиданному геометрическому вопросу о ширине многоугольника, целым точкам на эллипсоидах, подсчёту числа «счастливых билетов».

ISBN 5-900916-19-7

Выпуск данного сборника поддержан грантом
Российского фонда фундаментальных исследований
(номер проекта 96-01-14087)

Содержание

Математический мир

Умер ли Никола Бурбаки?

(Домыслы, легенды, достоверные и не очень достоверные сведения об одной несуществующей личности)

А. Б. Сосинский

В августе 1997 года известный французский математик Пьер Картье опубликовал, в виде препринта IHES[1)], материал под названием «Жизнь и смерть Никола Бурбаки», основная часть которого — интервью под заголовком «Продолжающий молчание Бурбаки», данное Картье 18.6.1997 г. журналистке М. Секстилл. Этот текст (который в расширенном виде опубликован в журнале «The Mathematical Intelligencer», V. 20, №1, 1998, с. 22–28) наверняка станет предметом живых обсуждений среди математиков: уж больно ярка и противоречива сама фигура математика Н. Бурбаки. Хотя бы потому, что (как знают теперь многие) такого учёного вообще не существует. Математик Бурбаки — фикция, что, впрочем, не помешало ему обрасти бóльшим числом легенд и сплетен, чем, скажем, Архимеду или Гауссу. Но давайте обо всём по порядку.

1. Взлёт: от студенческих шуток до мировой славы

В 1935 году группа молодых французских математиков из престижной парижской Эколь Нормаль Сюперьёр, лидерами которых стали Андре Вейль, Жан Дельсарт, Жан Дьёдонне, Анри Картан и Клод Шевалле, недовольные тем, как тогда преподавалась математика во Франции, взялись за пересмотр всей этой науки, от самых её оснований. Люди они

[1)] IHES — аббревиатура Института Высших Научных Исследований, расположенного под Парижем; это небольшой, но элитный математический центр, в котором Картье является одним из шести постоянных профессоров.

были не только молодые, со свойственным молодостью максимализмом, но и фрондёрски настроенные, с безжалостным чувством юмора и презрением к признанным авторитетам, особенно академического толка[2)].

Они организовали, сначала полушутливым образом, тайное сообщество — коллективного автора будущих трактатов, выпускаемых под единым псевдонимом (членство в нем впоследствии часто бывало секретом Полишинеля, но тем не менее этот секрет рьяно охранялся — но об этом ниже). Выбор псевдонима, отчасти случайный, отражал их (тогда шутливый) дух. По одной из версий, на выбор их подтолкнула конная помпезная статуя генерала Н. Бурбаки, маячившая на площади города Нанси перед кафе, где тогда группа подолгу обсуждала будущие книги; генерал прославился вовсе не военными успехами, а удивительной стратегической и тактической бездарностью и глупостью (проявившейся, в частности, в Крымской кампании в России), его фамилия (на самом деле франко-греческая) с итальянским окончанием „i“ (или японским: „aki“?) маскировала его реальное происхождение; впоследствии этот интернационализм был усугублен присуждением Н. Бурбаки звания профессора (несуществующего!) Университета Нанкаго[3)]. Впрочем, в достоверности этой версии есть сомнения (как и во многих других легендах о Бурбаки): при недавнем посещении Нанси автор этих строк никакой конной статуи генерала там не обнаружил.

За «обновление всей математики» бурбакисты взялись с рьяным энтузиазмом молодости[4)]. Довольно быстро были определены

1) главная цель трактата: «дать прочные основания всей современной математики в целом»;

2) общие принципы: единство и полная формализация математики на основе теории множеств; систематичность; догматизм и самодостаточность; изложение всегда идущее от общего к частному; ключевая роль понятия «структуры»;

[2)] П. Картье отмечает, что бóльшая часть членов группы тогда и впоследствии была протестантского или иудейского вероисповедания (в то время как почти все французы — католики); с этим он связывает как их фрондёрские настроения, так и нелюбовь к графическим изображениям (а, значит, и геометрии), не допускаемым в храмы протестантами и иудеями.

[3)] Nancago = Nancy + Chicago; в Чикаго работал член Бурбаки следующего призыва Сэм Эйленберг, туда же уехал в первые военные годы А. Вейль.

[4)] Бурбаки, пока он жив, всегда молод (по определению!): устав предписывает автоматическое исключение любого члена по достижению 50-летнего возраста; впрочем, исключение может произойти и раньше — но об этом ниже.

3) общее название всего трактата (*Начала математики*[5)], как у Евклида);

4) оглавление и список ключевых понятий;

5) структура трактата, название томов и выпусков.

Уже в 1939 году в издательстве «Эрманн» стали появляться первые выпуски. Работа не прекращалась (хотя и несколько замедлилась) в годы Второй Мировой Войны. Первые же послевоенные годы ознаменовались небывалым (а для бурбакистов совсем неожиданным) издательским успехом *Начал*. Пошли переводы на основные языки мира, а полунищее и полушутливое «секретное общество» вдруг оказалось всемирно знаменитым и (в меру) богатым.

2. Догматизм и формализация

Прежде чем продолжить наш рассказ о «жизни» Никола Бурбаки — несколько слов о его творчестве. Ключевые слова, его характеризующие, это — абстрактность, формализация, систематичность, догматизм.

Первые выпуски содержали вкладыш, озаглавленный «Инструкция к употреблению данного трактата», начинающийся со слов: «Настоящий трактат излагает математику с самых её начал и даёт полные доказательства $\langle\cdots\rangle$ ставя себе целью развивать базовые понятия, появляющиеся в большинстве задач современной математики $\langle\cdots\rangle$ эти понятия представлены в максимальной общности, а значит и очень абстрактно...». Далее там можно прочесть: «Принятый способ изложения является аксиоматическим и абстрактным $\langle\cdots\rangle$ для этого совершенно необходимо сразу же вооружиться весьма большим числом общих понятий и принципов...».

Эти общие принципы и понятия доставляются математической логикой, теорией множеств и придуманным бурбаками понятием *структуры*; на их основе строится общая топология, теория вещественных и комплексных чисел, а лишь затем — дифференциальное и интегральное исчисление (при этом используются малоизвестные даже сейчас в России понятия фильтра и ультрафильтра). Более подробно с «Началами» Бурбаки читатель может познакомиться непосредственно по их русскому переводу; особого внимания заслуживает предисловие редактора перевода (В. А. Успенского), помещённое в томе *Теория Множеств*, а также перевод статьи Бурбаки «Архитектура математики» и комментарий к ней

[5)] По-французски — «Eléments de Mathématique»; именно так, а не «de Mathématiques»: единственное число подчёркивает единство всей математики.

А. А. Ляпунова, опубликованные во второй серии журнала «Математическое просвещение»[6]. Мы же вернемся к нашему жизнеописанию Бурбаки.

3. Эпоха расцвета

Коммерческий успех первых книг трактата Бурбаки не привёл к уменьшению работоспособности группы: напротив, в первые послевоенные годы работа пошла с удвоенным энтузиазмом. Этому способствовало вливание новой крови: к группе Бурбаки присоединились, в частности, Ж. Диксмье, Р. Годеман, Ж.-Л. Кошуль, П. Самюэль, Ж.-П. Серр, упомянутый ранее С. Эйленберг и Л. Шварц.

Улучшение финансовой ситуации в группе привело к некоторым изменениям в её стиле работы. Говорят (никакой достоверной информации, повторяю, нет), бурбакисты собирались на ежегодные сессии в уютной вилле на средиземноморском побережье для обсуждения содержания и конкретных текстов очередных томов и их переизданий. Собрания продолжали быть весёлыми, бурными и плодотворными; по мере роста финансовых успехов улучшалось качество еды и вина и, до поры до времени, эффективность работы. В середине пятидесятых к группе присоединилась «третья волна»: А. Борель, Ф. Брюа, А. Гротендик, П. Картье, С. Ленг и Дж. Тейт.

Идеи Бурбаки начинают побеждать в университетском образовании, во всяком случае во Франции. Члены группы постепенно завоевывают ключевые посты во французских университетах и в CNRS[7]. Из молодых фрондёров бурбакисты постепенно становятся ведущими фигурами французского математического истеблишмента. Хотя ведущие члены группы Бурбаки, как индивидуальные учёные, получают единодушное признание во всех ведущих математических странах, реакция там на книги Бурбаки весьма неоднозначна — от безразличия в Англии до довольно враждебного отношения в некоторых математических центрах в США и в Советском Союзе в целом.

Последнее не помешало публикации трактатов Бурбаки на русском языке в середине шестидесятых годов и любопытству к отдельным членам группы со стороны молодых советских математиков, особенно во время Всемирного Математического Конгресса, проходившего в Москве

[6] *Н. Бурбаки (Франция – США).* Архитектура математики (Перевод с французского Д. Н. Ленского) // Математическое просвещение, №5, 1960 г. С. 99–112.
А. А. Ляпунов. О фундаменте и стиле современной математики. (По поводу статьи Н. Бурбаки.) Там же. С. 112–116.

[7] CNRS — Национальный Центр Научных Исследований, что-то вроде нашей системы академических институтов.

в 1966 году. Тогда запомнились как экстравагантный образ «действующего» бурбакиста Адриена Дуади (выступавшего с секционным докладом босиком и в рваных джинсах), так и уже выбывший из группы по возрасту (но официально представляющий её коммерческие интересы) Жан Дьёдонне, уверенно пропагандирующий крайне бурбакистские взгляды на математическое образование (но об этом — ниже). Однако у тогдашней русскоязычной математической молодежи дело дальше любопытства не шло: «естественно-научные» традиции русской математики оказались слишком крепкими, чтобы формалистические взгляды автора новых *Начал* могли увлечь её.

Однако даже самые стойкие противники бурбакизма подспудно испытали его влияние. Во всяком случае стиль практически всех научных работ по математике в период от пятидесятых по семидесятые годы постепенно изменился в сторону формализации, стал в той или иной степени походить на формально-бурбакистскую манеру, притом, как правило, этот процесс происходил неосознанно.

4. Жестокий юмор: легенды

О периоде расцвета Бурбаки бытует много легенд. Я расскажу здесь лишь три истории, заранее оговорившись, что (хотя они почерпнуты из вполне солидных изданий) за их достоверность ручаться нельзя.

Урод Жана Дьёдонне. Один из самым трудных томов трактата Бурбаки — том, посвящённый интегрированию (мере Хаара). Жану Дьёдонне, самому рьяному критику очередной рукописи этого тома, в конце сороковых годов было поручено написание очередной версии. Дьёдонне, в то время находившийся в расцвете своих творческих сил, забросив на целый год всю свою собственную математическую работу, полностью отдался этому нелегкому труду.

Через год, к назначенному сроку, он привез на бурбаковскую виллу на Средиземном море 12 (по числу членов группы) экземпляров своего труда. Первое обсуждение произошло на следующий вечер. Сидели в удобных креслах в большой гостиной, потягивая (в то время ещё не самое лучшее) красное вино и глядя на уютно разгоревшийся камин. Выступали поочередно, причем тон выступлений, неожиданно для гордившегося своим детищем Дьёдонне, был резко критическим. Первое же выступление завершилось такой оценкой: «Место этому уроду — здесь!», после чего помятые машинописные листки рукописи, с тщательно вставленными от руки формулами, были отправлены в камин. И так завершились все одиннадцать выступлений. Обиженный Дьёдонне удалился в свою комнату,

где на письменном столе к счастью оставался последний, его собственный, экземпляр рукописи.

Можно представить себе его ужас, когда вместо рукописи он обнаружил там лишь маленькую кучку пепла и записку: «Здесь покоится прах последнего урода Дьёдонне.»

Расставание с Андре Вейлем. В 1956 А. Вейлю, одному из основателей и бесспорных лидеров группы Бурбаки, должно было исполниться 50 лет; это означало, что близился срок его автоматического исключения из группы. Однако до него стали доходить слухи о том, что «молодые волки», недавно пополнившие группу, настроены его изгнать и раньше: это можно было сделать в полном соответствии с уставом, предусматривающим исключение за «профессиональную некомпетентность». Трудно было Вейлю держаться в курсе всех работ молодежи, среди которых уже царил А. Гротендик; но Вейль очень старался: уж очень хотелось ему избежать подобного позора. Поэтому за выступлением (очень интересным, но путаным) одного из молодых на семинаре Бурбаки он следил с неотступным вниманием, часто прерывая докладчика умными вопросами, радуясь, что он, пятидесятилетний старик Вейль, не теряет нить доклада, в то время как более молодые уже совсем запутались и даже перестали слушать.

Бедный Вейль! Когда доклад закончился, он узнал , что оказался жертвой тщательно отрежиссированного розыгрыша: последние 15 минут докладчик (с ведома всех слушателей, кроме Вейля) нес бессмысленную ахинею! Таким образом, за два месяца до своего пятидесятилетия один из отцов–основателей группы Бурбаки был изгнан из её рядов за профнепригодность. Следует отметить, что здесь вовсе не произошло сведения счётов, связанного с внутренними распрями (распри стали появляться лишь несколько позже) — Андре Вейля любили (в том числе и молодежь) — просто такое уж чувство юмора было у Никола Бурбаки.

Х. Боас и «коллективный псевдоним». В начале пятидесятых годов в США была опубликована математическая энциклопедия. В её издании принял активное участие тогда ещё молодой американский математик Харольд Боас, которому, в частности, была поручена статья *Никола Бурбаки*. Он написал: «Н. Бурбаки — коллективный псевдоним группы молодых французских математиков, занимающихся издательской деятельностью и . . . ». Через несколько дней после выхода книги в свет он получил лаконичное письмо: «Вас ждёт страшная кара. Н. Бурбаки».

И действительно, легко представить, в каком шоке был бедный Боас, когда через пару месяцев прочитал в реферативном журнале следующую рецензию на свою очередную работу: «Х. Боас — коллективный

псевдоним группы молодых американских математиков, занимающихся издательской деятельностью. В работе исследуется $\langle\cdots\rangle$, однако сформулированные результаты малозначительны, к тому же имеется грубая ошибка в ключевой Лемме 3.2 ...».

В конце рецензии стояла подпись: «Н. Бурбаки (Университет Нанкаго)». Добавлю от себя, что через несколько лет один молодой французский коллега искренне жаловался мне, как разные некомпетентные люди, в частности в Америке, пытаются подражать Бурбаки.

5. Бурбакизация образования

Мы уже отмечали влияние трактата Бурбаки на стиль изложения математических работ во всём мире и использование некоторых его томов как учебников во французских университетах. Однако — вопреки разумному сопротивлению многих математиков и педагогов — бурбакизму было суждено оказать существенное влияние и на школьное математическое образование во всём мире. В шестидесятые годы, к ужасу учителей и родителей школьников, стиль Бурбаки ворвался во все учебники, пошла волна увлечения «новой математикой». Не столько теория множеств, сколько полная формализация на основе понятия *алгебраической структуры* ставилась бурбакистами во главу угла школьного курса, из которого фактически была изгнана вся настоящая геометрия.

Опыт такого преподавания, особенно в крайних своих формах (например, в Бельгии и во Франции), сам и продемонстрировал свою несостоятельность. В семидесятых-восьмидесятых годах во всём мире стал наблюдаться постепенный откат от бурбакистских концепций. Однако целому поколению школьников прививалось совершенно одностороннее представление о математике как о науке, занимающейся формально-логическим преобразованием одной нудной тавтологии в другую. Меньше всего от школьного бурбакизма пострадала Великобритания (в основном проигнорировавшая его) и Россия, где реформа, возглавляемая А. Н. Колмогоровым, не оказалась под сильным влиянием Бурбаки и где замечательные традиции кружков и олимпиад позволили сохранить увлечённость математикой среди лучших школьников[8)].

[8)] Не вступая в полемику по поводу положительных и отрицательных сторон колмогоровской реформы, считаю необходимым отметить следующее. Резкая (и, к сожалению, в итоге успешная) критика его программы и общих концепций, в частности в выступлениях Л. С. Понтрягина, была во многом основана на искажении фактов и порой на квасном патриотизме. А. Н. Колмогоров в первую очередь обвинялся в «бурбакизации школьного образования на основе чуждого русской математике теоретико-множественного подхода». Однако источником реформы на самом деле было вовсе не

6. Упадок

Тогда, когда преуспевающая группа Бурбаки была на самом гребне успеха, когда её слава распространилась по всему миру, когда развивалось её влияние на все университеты и школы и когда лучшие математики Франции (а иногда и других стран) пополняли её ряды, в ней самой уже зрели причины надвигающегося упадка. Одна из главных — однобокость. Хотя лидеры Бурбаки всегда были математиками-универсалами, они практически все были ближе к алгебре, чем к другим разделам математики. Не было среди них ни геометров, ни настоящих аналитиков[9]: интегральное и дифференциальное исчисление они воспринимали как раздел функционального анализа, как раз в то время, когда происходила постепенная геометризация анализа на основе современной дифференциальной геометрии и топологии. Они не увидели (или не захотели увидеть и отразить в своем трактате) начавшееся в семидесятые годы (на той же основе) слияние теоретической физики и математики, в частности вокруг квантовой механики. И если, в известной степени, Никола Бурбаки удалось завершить свой замысел обоснования всей математики на единой основе (в пяти томах «Фундаментальных структур»), его дальнейшие публикации очень однобоко отражали современную математику, не сумели уловить ни дух, ни направления её магистрального развития.

Внутри группы начались склоки. Великий Гротендик разругался с другими членами сообщества (в том числе и со своими учениками) и покинул Бурбаки. Его авторитет, в период его членства в Бурбаки и позже, лишь усилил дух алгебраической формализации, но так и не вылился в появление новых томов трактата. Другие крупные математики (Жак Титс, Серж Ленг) ушли ещё быстрее, тоже по собственной инициативе, а зубры первых призывов постепенно выбывали по возрасту. Пополнение

влияние Бурбаки, а просто объективная оценка недопустимого отставания школьной математики от математики нашего времени, необходимость введения в среднюю школу важнейших (в частности для приложений) математических понятий: геометрические преобразования, производная и интеграл, алгоритмика, комбинаторные понятия теории множеств. По иронии судьбы противник педагогических концепций Бурбаки А. Н. Колмогоров обвинялся именно в бурбакизации образования, в заполонении школьной математики формалистикой теории множеств (в то время как на самом деле этой теории в программе вовсе не было, а присутствовали в умеренных дозах лишь некоторые её понятия). В целом можно сказать, что реформа школьной математики шестидесятых годов в России оказалась заметно умереннее и разумнее, чем в Западной Европе или США.

[9] Характерно, что ни геометрический тополог Рене Том, ни тополог-аналитик Жан Лере не входили (насколько мне известно) в группу Бурбаки.

группы происходило за счёт математиков совсем другого, более мелкого, калибра.

Между Бурбаки и его издательством «Эрманн» начались и финансовые распри, которые стали отнимать больше сил и времени, чем написание трактатов. Когда Бурбаки (чьи финансовые интересы представлял уже покинувший группу по возрасту Картье) наконец выиграл затянувшийся судебный процесс у «Эрманна» в 1980 году и перешёл в другое издательство, новые тома уже перестали создаваться. В 1983 году появилась последняя новая (вернее обновлённая) публикация Никола Бурбаки. Ему было 48 лет, т. е. оставалось всего два года до рокового пятидесятилетнего возраста, им же учреждённого. В одном из учёных журналов появился его некролог, выдержанный в тех традициях жестокого юмора, о которых говорилось выше. Я помню, как все смеялись (это — над некрологом!), когда Юрий Иванович Манин зачитал его на очередном заседании Московского Математического Общества.

7. Кончина?

Но умер ли Никола Бурбаки? И да, и нет. Как юридическое лицо он ещё существует. Существует и замечательный *Семинар Бурбаки*, и публикуемый им журнал; на семинаре выступают ведущие математики мира (не только члены группы) с обзорными докладами о крупнейших достижениях современной математики (при этом, что очень разумно, доклады, как правило, не поручаются самим авторам этих достижений). Но как творческий математик, и даже как великий методист, Бурбаки умер. Возродится ли он, чтобы спеть нам свою лебединую песню? Думаю — вряд ли.

Почему? Бурбакизация французского среднего и высшего образования привела к тому, что выпускники вузов Франции, в частности, Эколь Нормаль Сюперьёр (из которых, в основном, и пополнялась группа), постепенно превращались в карикатуры схоласта Бурбаки, погрязнув в абстрактных алгебраических схемах, оторванных от живого развития математики. Эпохе великого поколения математиков Франции, эпохе А. Вейля, А. Картана, К. Шевалле, Ж.-П. Серра, Л. Шварца, П. Картье, А. Бореля, П. Делиня и А. Гротендика пришёл конец. Пришедшие им на смену математики, обученные в школе и вузе по схоластическим принципам бурбакизма, оказались другого калибра: Никола Бурбаки сам себя и уничтожил.

Мы хотели бы постоянно обсуждать тему «Учитель и ученик» в нашем журнале. В отечественной науке, как, пожалуй, нигде еще, практически для каждого крупного математика на начальном этапе его карьеры огромную роль играет общение с Учителем, который поставил первую задачу, был первым, кто выслушал её решение, ободрил и подсказал, чем заниматься дальше.

В июне прошлого года исполнилось 60 лет одному из крупнейших математиков современности, главе блестящей научной школы — академику Владимиру Игоревичу Арнольду. К его юбилею московское издательство «Фазис» выпустило замечательную книгу «Владимир Игоревич Арнольд. Избранное–60». Редакционную подготовку материалов осуществлял один из учеников В. И. Арнольда — Михаил Борисович Севрюк. Мы предлагаем читателям фрагмент из интервью, данного им В. М. Тихомирову для нашего журнала. Оно вполне соответствует теме «Учитель и ученик». Мы надеемся, что читателю будет интересно узнать, как постепенно В. И. Арнольд вводит своего ученика в науку, как экзаменует, как и какие ставит задачи и большие проблемы.

Мой научный руководитель — В. И. Арнольд

М. Б. Севрюк

Огромный вклад, внесённый Владимиром Игоревичем в математику второй половины уходящего века, относится ко всем сторонам научного творчества — и к решению задач, причем задач классических, которые стояли десятилетиями, и к постановке новых проблем, и к созданию новых теорий, и к совершенствованию математического образования.

Говоря о классических задачах, достаточно упомянуть решённую им тринадцатую проблему Гильберта. Владимир Игоревич ещё в студенческие годы сделал (вслед за прорывом в этой области, совершенным его учителем Андреем Николаевичем Колмогоровым) завершающий шаг в доказательстве того, что любую непрерывную функцию трёх переменных можно представить как суперпозицию непрерывных функций двух переменных. Другой пример — решение классической проблемы Биркгофа об устойчивости эллиптической неподвижной точки отображения плоскости на себя, сохраняющего площадь. Третий пример — доказательство существования большого числа квазипериодических движений в планетных системах (проблема, стоявшая в небесной механике, по крайней мере, со времен Пуанкаре).

В том, что касается построения теорий, необходимо сказать, что Владимир Игоревич открыл много новых путей в математике и нашёл очень

много неожиданных связей между различными её областями. Большинство новых математических теорий, созданных им, основывается именно на таких связях. Он сам писал в одной из своих последних работ «От суперпозиций до теории КАМ» (мемуарного характера), что обнаружение связей между, казалось бы, совершенно далекими друг от друга вещами — это одно из самых больших наслаждений, которое может дать математику наша наука, и ему выпало счастье испытать это наслаждение несколько раз.

Например, в теории особенностей (а Владимир Игоревич является одним из её создателей) им была открыта фундаментальная связь между критическими точками гладких функций и группами Кокстера, что привело к современной классификации особенностей. Ему принадлежит установление связи между шестнадцатой проблемой Гильберта об овалах вещественных алгебраических кривых и четырёхмерной топологией. Это вызвало прорыв в исследованиях, связанных с шестнадцатой проблемой Гильберта. Владимир Игоревич нашёл совершенно замечательную связь между теорией кос, с одной стороны, и теорией особенностей и алгебраической геометрией, с другой. И во всех этих случаях речь идёт о построении новых обширных математических теорий, а иногда о совершенно новом взгляде на теории, существовавшие ранее.

Ему принадлежит также создание теорий несколько иного рода — когда в задаче, включающей в себя очень много различных структур, и алгебраических, и топологических, Владимир Игоревич устанавливал, какие именно структуры ответственны за тот или иной эффект, и на основе подобного тщательного анализа обобщал известный результат в самых разных направлениях. Знаменитым примером является создание симплектической топологии, которое началось с его статьи 1965 г. в C. R. Acad. Sci. Paris. В этой работе и в ряде последующих Владимир Игоревич выдвинул ряд гипотез, связанных с числом неподвижных точек симплектоморфизмов, и в дальнейшем он возвращался не раз к этой тематике в течение всей своей математической карьеры. В последние годы им обсуждаются псевдопериодическая топология, проективная топология, градиентная топология...

Что же касается постановки новых задач, то я хотел бы сказать следующее. Каждый семестр первое заседание семинара по теории особенностей (на мехмате МГУ), которым Владимир Игоревич руководит на протяжении уже очень многих лет, он посвящает именно постановке новых задач, которые подхватываются его учениками, участниками семинара, и из этих задач часто впоследствии возникают целые новые направления. Творчество Владимира Игоревича очень обширно и охватывает самые

разные разделы математики, механики и физики — от теории дифференциальных уравнений до теории чисел, от математической логики до алгебраической геометрии, от топологии до гидродинамики, от теории катастроф до космологии, и во многих областях науки его результаты являются основополагающими. В книге «Владимир Игоревич Арнольд. Избранное–60» приведены составленный самим автором полный список его основных результатов и основная тематика его исследований.

Наконец, говоря о вкладе Владимира Игоревича в математическое образование, мне хотелось бы процитировать отрывок из рецензии на третье издание его знаменитого учебника «Математические методы классической механики» в Math. Reviews, MR 93c:70001 (рецензент A. Iacob):

«Пусть $S_1 = \{$наиболее влиятельные книги второй половины XX века$\}$, $S_2 = \{$наиболее часто цитируемые книги$\}$, $S_3 = \{$книги, имеющие наибольшую вероятность сохранить свою актуальность в XXI веке$\}$, $S_4 = \{$книги, очень полезные в преподавании$\}$, $S_5 = \{$книги, написанные в совершенно необычном, присущем только данному автору стиле$\}$, $S_6 = \{$книги, читать которые доставляет подлинное наслаждение$\}$, $A =$ рецензируемая книга. Предложение: $A \in \cap_{i=1}^{6} S_i$.»

Мне кажется, что это предложение справедливо для всех книг Владимира Игоревича.

Я стал заниматься у Владимира Игоревича по совету многих людей, стал ходить на его семинары. На первом курсе я с ним переговорил и он дал мне список задач. Я ни одной из этих задач не решил, и когда я вернулся к нему совершенно обескураженный, он сказал, что в принципе этого и ожидал. Но затем, когда он мне дал ряд задач на лето между первым и вторым курсами (само собой разумеется, это были известные задачи), я тогда решил всё, причем одну из этих задач (я не помню, в чем она заключалась) я решил совершенно не тем способом, который он ожидал, хотя сама по себе задача никоим образом не была ни новой, ни сложной.

Сначала Владимир Игоревич предложил мне заниматься вопросом о числе нулей абелевых интегралов — одной из составных частей шестнадцатой проблемы Гильберта. Эта задача связана с числом предельных циклов, которые возникают при малом негамильтоновом возмущении гамильтоновой системы с одной степенью свободы (или, более общим образом, при малом возмущении векторного поля на плоскости, имеющего первый интеграл). В этой задаче у меня никаких существенных продвижений не было. А так как я с Владимиром Игоревичем стал заниматься очень рано, то к тому моменту, когда надо было писать курсовую

работу на третьем курсе, ещё было время сменить тематику. И тогда он привлёк меня (видя, что в общем-то у меня продвижения нет) к задаче о представимости алгебраических функций большого числа переменных в виде суперпозиции функций меньшего числа переменных. В качестве ключа к получению новых серьёзных результатов в этой области Владимир Игоревич предложил мне вычислить различные топологические характеристики (в частности, целочисленные гомологии и когомологии) некоторых специальных алгебраических подмножеств пространства комплексных бинарных форм произвольной фиксированной степени. Основное внимание уделялось таким довольно мощным средствам, как теория Ходжа, — предполагалось, что важную роль в дальнейших исследованиях вопросов представимости будут играть так называемые смешанные структуры Ходжа на рациональных когомологиях подходящих алгебраических подмножеств пространства бинарных форм.

Мне удалось получить ряд интересных результатов о гомологиях таких подмножеств. Я написал очень хорошую курсовую работу на третьем курсе, и на следующий год моя курсовая была снова посвящена этой тематике. Результаты этих работ составили содержание двух заметок, одной — в «Успехах математических наук», другой — в «Вестнике Московского университета». Но надежды на продвижение в первоначальной задаче о суперпозициях не оправдались, потому что вычисленные мной смешанные структуры Ходжа оказались слишком простыми. Тогда Владимир Игоревич привлёк меня к совершенно новой тематике — к динамическим системам, которыми я и занимаюсь с тех пор вплоть до настоящего времени.

Сначала я просто сдал экзамен по спецкурсу, который читал Владимир Игоревич — это было во втором семестре четвёртого курса. Спецкурс был по обыкновенным дифференциальным уравнениям, я сдавал экзамен по этому курсу как один из экзаменов, которые нужно было сдавать в соответствии с учебным планом. Но мне в качестве экзамена (а сдача экзаменов по спецкурсам Владимира Игоревича всегда заключалась исключительно в решении задач — разумеется, не обязательно новых) была дана интересная задача о свойствах пар (A, G) диффеоморфизмов плоскости с общей неподвижной точкой, удовлетворяющих соотношению $AGA = G$ (под произведением отображений здесь понимается их композиция).

Я сдавал Владимиру Игоревичу два спецкурса — по теории особенностей и по дополнительным главам обыкновенных дифференциальных уравнений. В качестве задачи по теории особенностей он предложил мне доказать одну классификационную теорему, принадлежащую М. Джу-

сти и приведённую без доказательства в монографии В. И. Арнольда, А. Н. Варченко и С. М. Гусейн-Заде «Особенности дифференцируемых отображений I» (см. стр. 131). Я, конечно, всё это сделал. Задачу про диффеоморфизмы плоскости я также успешно решил (на уровне формальных степенных рядов). Не помню, до какой степени полученные результаты были новыми, но после этой задачи я стал заниматься *обратимыми системами*.

Обратимые системы — это замечательный класс динамических систем, инвариантных относительно одновременного применения двух операторов — обращения времени и некоторого диффеоморфизма фазового пространства (если два диффеоморфизма A и G связаны соотношением $AGA = G$, т. е. $GAG^{-1} = A^{-1}$, то динамическая система с дискретным временем, порождённая диффеоморфизмом A, обратима относительно преобразования G). И вот начиная с задачи, которая на четвёртом курсе была предложена мне Владимиром Игоревичем просто в качестве задачи при сдаче спецкурса, мои занятия обратимыми системами привели к получению достаточно большого числа результатов, сделавших меня одним из лидеров этого направления, и составили существенную часть как моей математической деятельности, так и моей жизни в целом. Владимир Игоревич предложил мне заниматься обратимыми системами прежде всего применительно к той теории, в которой он является одним из основателей — теории КАМ (Колмогорова–Арнольда–Мозера).

Основателем теории обратимых систем по праву нужно считать Дж. Д. Биркгофа, который исследовал такие системы ещё в начале века. К тому моменту, когда я стал заниматься обратимыми системами и теорией КАМ, работ по обратимым системам было ещё относительно немного. С другой стороны, в ряде случаев, когда Владимир Игоревич ставит перед учеником новую задачу, он обрисовывает её исключительно на уровне идей. Он не даёт своему ученику ни списка основных публикаций, которые уже существуют в данной области, ни списка результатов или тех людей, которые занимались данными вопросами, не предлагает ученику ознакомиться с какой-либо вводной литературой по данной тематике. Он просто рассказывает на чисто идейном уровне, какие проблемы существуют в данной области, почти без указания каких-либо конкретных имен и публикаций, и предполагает, что ученик сам должен всё это находить. Такой подход часто оказывается чрезвычайно плодотворным — не только с той точки зрения, что он развивает в ученике самостоятельность, но также и потому, что позволяет иногда человеку, только начинающему заниматься данной тематикой, получать такие результаты, которые вряд ли могли быть им найдены, если бы этот человек до

начала самостоятельной работы основательным образом проштудировал работы своих предшественников. Так произошло и со мной.

Я сразу стал пытаться (практически ещё ничего не зная об обратимых системах и не имея никакого опыта в теории КАМ) доказывать теорему КАМ для обратимых систем без стандартного предположения о том, что обращающий диффеоморфизм фазового пространства является инволюцией (отображением, квадрат которого есть тождественное преобразование), и мне удалось получить обратимую теорему КАМ без такого предположения! Я и по настоящее время чрезвычайно горжусь этим достижением. Результат, который и по сей день кажется мне совершенно удивительным и замечательным, заключается в следующем. Предположим, что у нас есть интегрируемая обратимая система. Ее фазовое пространство расслоено на инвариантные торы, движение по которым в подходящих угловых координатах происходит с постоянной скоростью. Обращающий диффеоморфизм заключается в изменении знака угловых координат на этих торах и является инволюцией. Теперь предположим, что мы эту систему чуть пошевелим, не выводя её из класса обратимых, но при этом будем возмущать не только саму систему, но и обращающий диффеоморфизм, *не предполагая, что возмущённый обращающий диффеоморфизм по-прежнему является инволюцией.* Тогда, если исходная система была невырожденной в том смысле, что частоты движения на невозмущённых инвариантных торах невырожденным образом зависели от параметра, нумерующего торы, то в аналитической категории возмущённая система будет иметь много инвариантных торов (близких к невозмущённым торам и инвариантных также относительно обращающего диффеоморфизма), движение по этим торам будет квазипериодическим... — т. е. будет иметь место стандартная ситуация обычной (гамильтоновой) теории КАМ, но при этом, кроме того, возмущённый обращающий диффеоморфизм *обязательно будет являться инволюцией.* Итак, возмущая невырожденную интегрируемую обратимую систему, мы не выходим за рамки обратимых систем с инволютивным обращающим диффеоморфизмом, но этого, оказывается, не нужно требовать a priori — это получается само собой *(жёсткость инволютивности).*

Эта теорема, аналогичное локальное утверждение и другие результаты составили содержание моей дипломной работы, а затем вошли в книгу, которую я под непосредственным наблюдением Владимира Игоревича написал во время обучения в аспирантуре. Она была опубликована в серии Lecture Notes in Mathematics. И после окончания аспирантуры и защиты диссертации мне много раз выпадала счастливая возможность принимать щедрую помощь и поддержку Владимира Игоревича.

О филдсовских медалях

М. И. Монастырский

Пожалуй, самыми известными среди международных премий, которых удостаиваются математики, являются *премии Филдса*. Основатель этой премии — Джон Чарльз Филдс (1863 – 1932). Он родился в Канаде, окончил университет Торонто, где и работал (после длительного пребывания в Европе) с 1902 г. до конца жизни. Научные работы Филдса были связаны с теорией алгебраических функций и алгеброй. Но наибольшую известность имя Филдса получило в связи с его общественной деятельностью. На Международном математическом конгрессе в Канаде (Торонто, 1924) впервые обсуждалась его идея организации международной премии. Филдс составил меморандум, в котором охарактеризовал статут новой премии. Он писал: «Я особо подчёркиваю, что медаль должна быть интернациональна и объективна, насколько это возможно. $\langle\cdots\rangle$ Она ни под каким видом не должна включать упоминание о какой-либо стране, институте или личности». Идея Филдса осуществилась на международном конгрессе в Осло (1936). И с тех пор медали выдающимся математикам вручались на всех международных математических конгрессах. На этих медалях (в отличие от нобелевских) гравируется лишь фамилия лауреата и год присуждения премии. И хотя там нет никакого упоминания о Филдсе, и за премией, и за медалью заслуженно закрепилось его имя.

В меморандуме Филдса говорилось, что премия должна не только отмечать уже достигнутые результаты, но и стимулировать дальнейшую творческую активность лауреата. Первым составом Филдсовского комитета эта фраза была истолкована, как указание, что премии должны вручаться относительно молодым учёным. Чуть позже установилась возрастная граница: возраст лауреата не должен превосходить 40 лет.

Медалями Филдса были награждены следующие математики (указано место работы лауреата к моменту присуждения медали):

1936 год. Д. Дуглас (1897–1965) (Массачусетский технологический институт, США); Л. Альфорс (1907–1996) (Хельсинский университет, Финляндия).

1950 год. Л. Шварц (1915) (Университет Нанси, Франция); А. Сельберг (1917) (Институт высших исследований, Принстон, США).

1954 год. Ж.-П. Серр (1926) (Парижский университет, Франция); К. Кодаира (1915) (Принстонский университет, США).

1958 год. К. Ф. Рот (1925) (Лондонский университет, Великобритания); Р. Том (1923) (Университет Страсбурга, Франция).

1962 год. Л. Хёрмандер (1931) (Стокгольмский университет, Швеция); Дж. Милнор (1931) (Принстонский университет, США).

1966 год. С. Смейл (1930) (Калифорнийский университет, США); П. Коэн (1934) (Стэнфордский университет, США); А. Гротендик (1928) (Парижский университет, Франция); М. Атья (1929) (Оксфордский университет, Великобритания).

1970 год. А. Бейкер (1939) (Кэмбриджский университет, Великобритания); С. П. Новиков (1938) (Математический институт им. В. А. Стеклова, Москва, СССР); Д. Томпсон (1932) (Кэмбриджский университет, Великобритания); Х. Хиронака (1931) (Гарвардский университет, США).

1974 год. Д. Мамфорд (1937) (Гарвардский университет, США); Э. Бомбиери (1940) (Пизанский университет, Италия).

1978 год. П. Делинь (1944) (Институт высших исследований Бюрсюр-Иветт, Франция); Д. Квиллен (1940) (Массачусетский технологический институт, США); Г. А. Маргулис (1946) (Институт проблем передачи информации, Москва, СССР); Ч. Фефферман (1949) (Принстонский университет, США).

1983 год. А. Конн (1947) (Парижский университет, Франция); В. П. Тёрстон (1946) (Принстонский университет, США); Ч. Т. Яо (1949) (Стэнфордский университет, США).

1986 год. С. К. Дональдсон (1957) (Оксфордский университет, Великобритания); Г. Фалтингс (1954) (Принстонский университет, США); М. Фридман (1951) (Университет Сан-Диего, Калифорния, США).

1990 год. В. Г. Дринфельд (1954) (Физико-технический институт низких температур, Харьков, СССР); Э. Виттен (1951) (Институт высших исследований, Принстон, США); В. Ф. Р. Джонс (Калифорнийский университет, Беркли, США); Ш. Мори (Университет Киото, Япония).

1994 год. Ж. Бургэн (1954) (Институт высших исследований, Париж, Франция); П. Л. Лионс (1956) (Университет Пари-Дофин, Франция); Ж. Х. Йоккос (1957) (Университет Пари-Зюд, Орсей, Франция); Е. Зельманов (1955) (Университет Висконсин, Медисон, США).

Более подробно обо всех этих лауреатах можно прочитать в книгах: *Монастырский М.* Премия Филдса. Знание. Математика/Кибернетика, 2, 1991; *Monastyrsky M.* Modern Mathematicsin the Light of Fields Medals. A. K. Peters, 1996.

А в этом номере в статье В. В. Успенского и В. М. Тихомирова рассказывается о замечательных результатах А. Н. Колмогорова и Л. С. Понтрягина, за которые они вполне могли бы получить филдсовские премии, но этого не произошло из-за железного занавеса (см. стр. 21–40).

Первые филдсовские лауреаты и советская математика 30-х годов. I

В. М. Тихомиров В. В. Успенский

Первые лауреаты филдсовской медали были названы на конгрессе в Осло в 1936 году. В состав филдсовского комитета, который принимал решение об этом присуждении, вошли крупнейшие математики того времени: итальянец Ф. Севери (председатель), американец Дж. Биркгоф, француз Э. Картан, японец Т. Такаги, К. Каратеодори (немецкий математик греческого происхождения).

Золотые медали и денежный приз (1500 долларов) были вручены двум математикам: Дж. Дугласу из Массачусетского технологического института (MIT) и Л. Альфорсу из Хельсинского университета.

Джесс Дуглас (1897 – 1965) — американский математик, работавший в Нью-Йорке и MIT. Результатов, сопоставимых с тем, за который он получил свою медаль, в его биографии больше не было. Дуглас не приехал на конгресс, сославшись на трудности длительного путешествия. Медаль была вручена Норберту Винеру, работавшему в MIT.

Ларс Альфорс (1907 – 1996) — финский математик. Он прожил большую жизнь, работал в Швейцарии, а с 1946 года — в США. В 1986 году, в год пятидесятилетия премии, Альфорс был избран почётным президентом конгресса в Беркли. Он выступил на конгрессе с воспоминаниями о первом присуждении.

Оба лауреата были представителями анализа. Дуглас получил премию за решение задачи Плато. Задача Плато состоит в доказательстве существования поверхности минимальной площади с заданной границей. Впервые она была поставлена Лагранжем в 1760 году. Бельгийский физик Плато показал (в 1849 г.), что минимальные поверхности могут быть получены в виде мыльных пленок, натянутых на проволочный каркас. Вопрос о существовании для любой жордановой кривой Г, расположенной в n-мерном (в частности, в трёхмерном) пространстве, поверхности минимальной площади с границей Г, стали называть задачей Плато. Задача Плато была почти одновременно решена двумя математиками: Д. Дугласом и Т. Радо. Решение Дугласа было сочтено достойным премии Филдса.

О теореме Дугласа – Радо и о многомерных обобщениях задачи Плато можно прочесть в [10, § 47, 48].

Альфорс был удостоен медали за развитие теории римановых поверхностей и разработку теории квазиконформных отображений. С тех пор в течение полувека Альфорс был одним из лидеров в комплексном анализе.

Ну а что же советские математики?

Трудно оспорить, что в 1936 г. советская математическая школа была самой выдающейся во всём мире. Нацисты разгромили немецкую школу, французская переживала период смены поколений, математическая школа США только набирала обороты. В нашей стране в тридцатые годы достигло расцвета творчество таких учёных (не старше 40 лет!), как П. С. Александров, А. О. Гельфонд, А. Н. Колмогоров, М. Г. Крейн, М. А. Лаврентьев, Л. А. Люстерник, Д. Е. Меньшов, П. С. Новиков, Л. С. Понтрягин, С. Л. Соболев, А. Я. Хинчин, Л. Г. Шнирельман — список можно продолжить. Ни в одной другой математической школе того времени не было такого соцветия выдающихся математиков!

Но железный занавес уже опустился. Контакты между нашей страной и остальным миром были прерваны. На конгресс в Осло были приглашены Гельфонд и Хинчин, но они не смогли поехать. Вообще, заграничные командировки закончились в 1931 году. Этим в значительной мере и объясняется то, что среди лауреатов 1936 года не было советских математиков.

А. Н. Колмогоров рассказывал, что когда однажды в середине 30-х годов известного американского математика С. Лефшеца спросили, кого из современных молодых математиков во всём мире он считает наиболее выдающимися, он назвал четыре имени: А. О. Гельфонд, А. Н. Колмогоров, Л. С. Понтрягин и Л. Г. Шнирельман.

Первое присуждение филдсовской медали было проведено по двум разным признакам: за решение проблемы (Дуглас) и создание теории (Альфорс). В творчестве четырёх названных советских математиков (и большинства других, перечисленных выше) были и решение замечательных проблем, и создание теорий, и разработка новых методов. В кратком очерке невозможно дать сколько-нибудь полное представление об их достижениях. Мы приведем по одному из наиболее эффектных результатов А. Н. Колмогорова и Л. С. Понтрягина. О результатах А. О. Гельфонда и Л. Г. Шнирельмана мы собираемся рассказать в другой статье.

Мы не следуем оригинальным доказательствам, а используем методические усовершенствования, накопленные за минувшие годы. Но основные идеи доказательств восходят к авторским работам.

Пример Колмогорова всюду расходящегося ряда Фурье

Андрей Николаевич Колмогоров (1903 – 1987) — один из величайших учёных двадцатого века, для которого была характерна необычайная широта творческих интересов. Вот неполный список тех областей математики, где он оставил непреходящий след: теория ортогональных рядов, дескриптивная теория множеств, математическая логика, классическая теория вероятностей, геометрия, случайные процессы, математическая статистика, функциональный анализ, теория приближений, топология, дифференциальные уравнения, теория стрельбы, турбулентность, теория алгоритмов, динамические системы, классическая механика, метрическая теория функций, теория информации, алгоритмическая теория вероятностей. Существенную компоненту в его исследованиях составляют работы в области смежных наук: в физике, биологии, геологии, океанологии, метеорологии, кристаллографии, стиховедении и т. п. О личности Колмогорова и о его творчестве см. [9, 4, 11].

Поступив в 1920 г. в Московский университет, А. Н. Колмогоров становится учеником Н. Н. Лузина. В годы аспирантуры он получает выдающиеся результаты в теории функций, теории множеств и математической логике (в том числе и тот знаменитый результат, о котором мы расскажем ниже — пример расходящегося ряда Фурье), начинает свои исследования в теории вероятностей. В начале 30-х годов он создаёт теорию марковских процессов, завершая усилия Эйнштейна, Планка, Смолуховского и Винера, а затем создаёт аксиоматику теории вероятностей, ставшую частью образования любого математика. Эти выдающиеся достижения были получены А. Н. Колмогоровым до 1935 года, когда рассматривались первые присуждения филдсовских медалей.

В 1921 году восемнадцатилетний студент второго курса физико-математического факультета Московского университета Андрей Колмогоров построил пример *интегрируемой функции, ряд Фурье которой расходится почти всюду.* Это одно из самых замечательных достижений в теории тригонометрических рядов, которой посвятили свои исследования многие крупнейшие математики, начиная с Эйлера и Фурье. После примера Колмогорова сорок с лишним лет оставалось неизвестным, существует ли *непрерывная* функция с тем же свойством. Лишь в 1966 году Карлесон показал, что пример Колмогорова нельзя усилить: для всякой функции с интегрируемым квадратом ряд Фурье сходится почти всюду к самой функции (в виде гипотезы это сформулировал Н. Н. Лузин в 1915 году). «Почти всюду» означает *всюду, за исключением множества меры нуль.*

Множество на отрезке имеет меру нуль, если для любого $\varepsilon > 0$ его можно покрыть последовательностью интервалов с суммой длин $< \varepsilon$.

Впоследствии Колмогоров построил пример ряда, расходящегося всюду (оба результата содержатся в книге [4, статьи 1 и 11]). Мы построим далее именно такой пример.

Функция будет строиться на отрезке $[-\pi;\pi]$. Каждой интегрируемой функции f на $[-\pi;\pi]$ сопоставляется её *ряд Фурье*

$$\frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx),$$

где

$$a_n = \frac{1}{\pi}\int\limits_{-\pi}^{\pi} f(t)\cos nt\,dt, \quad b_n = \frac{1}{\pi}\int\limits_{-\pi}^{\pi} f(t)\sin nt\,dt.$$

Сумма $a_0/2 + \sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx)$ называется *n-той суммой Фурье* и обозначается $S_n(x,f)$. Если $f(x) = a_0/2 + \sum_{k=1}^{N}(a_k \cos kx + b_k \sin kx)$ — тригонометрический многочлен, то a_k и b_k — это коэффициенты Фурье функции f, так что $S_m(x,f) = a_0/2 + \sum_{k=1}^{m}(a_k \cos kx + b_k \sin kx)$ при $m < N$ и $S_m(x,f) = f$ при $m \geqslant N$. Ряд Фурье дифференцируемой функции сходится к ней самой, ряд Фурье непрерывной функции может расходиться на бесконечном множестве точек (которое, однако, имеет меру нуль).

Понятие *интегрируемой* по Лебегу функции нам не понадобится в полном объеме, нам будет достаточно следующего свойства: если последовательность $\{p_n\}_{n=1}^{\infty}$ непрерывных неотрицательных функций на отрезке такова, что $\sum_{n=1}^{\infty}\int p_n < \infty$, то ряд $\sum_{n=1}^{\infty} p_n(t)$ сходится почти всюду к интегрируемой функции f, при этом коэффициенты Фурье этой функции являются пределами коэффициентов Фурье частичных сумм ряда $\sum_{n=1}^{\infty} p_n$. Это следует из теоремы Леви и теоремы Лебега о мажорируемой сходимости [5].

Сформулируем ещё раз результат Колмогорова, который мы будем доказывать:

Существует интегрируемая функция, ряд Фурье которой расходится в каждой точке.

Наше построение складывается из нескольких шагов.

Предложение 1. *Для любого положительного числа a найдётся тригонометрический многочлен p_a, такой, что $|p_a(x)| \leqslant 1$ для всех x и при этом*

$$\max_n S_n(0,p_a) > a.$$

Это предложение было известно задолго до 1921 года.

Доказательство. Для $S_n(0,f)$ существует явное выражение

$$S_n(0,f)=\int\limits_{-\pi}^{\pi} D_n(x)f(x)dx,\quad D_n(x)=\frac{1}{2\pi}\frac{\sin(n+1/2)x}{\sin x/2}.$$

(Функцию D_n называют *ядром Дирихле* в честь знаменитого немецкого математика французского происхождения, внесшего большой вклад в теорию рядов Фурье; Дирихле, в частности, доказал сходимость ряда Фурье при ослабленных требованиях гладкости.) Мы имеем

$$\int\limits_{-\pi}^{\pi} |D_n(x)|dx\to\infty. \tag{1}$$

Формула (1) следует из выкладки:

$$\int\limits_{-\pi}^{\pi} |D_n(x)|dx=2\int\limits_{0}^{\pi} |D_n(x)|dx \overset{\sin x/2\leqslant x/2}{\geqslant}$$

$$\geqslant \frac{2}{\pi}\int\limits_{0}^{\pi}\frac{|\sin(n+1/2)x|}{x}dx \overset{t=(n+1/2)x}{=} \frac{2}{\pi}\int\limits_{0}^{(n+1/2)\pi}\frac{|\sin t|}{t}dt.$$

Последний интеграл оценивается суммой $C(1+1/2+1/3+\cdots+1/n)$, откуда и следует (1). Фиксируем n, для которого $\int\limits_{-\pi}^{\pi}|D_n(x)|dx>a$. Рассмотрим разрывную функцию $s(x)=\operatorname{sgn} D_n(x)$, где $\operatorname{sgn} x=x/|x|$. Её можно приблизить непрерывной кусочно-линейной функцией f так, чтобы $|f(x)|<1$ при всех x и чтобы интегралы

$$\int\limits_{-\pi}^{\pi} D_n(x)f(x)\,dx \text{ и } \int\limits_{-\pi}^{\pi} D_n(x)s(x)\,dx=\int\limits_{-\pi}^{\pi}|D_n(x)|\,dx$$

были сколь угодно близки. Поэтому первый интеграл тоже будет $>a$. По теореме Вейерштрасса всякую непрерывную функцию на отрезке $[-\pi;\pi]$, принимающую равные значения на концах, можно равномерно приблизить тригонометрическими многочленами. Применяя это к f, находим тригонометрический многочлен p, для которого $|p(x)|\leqslant 1$ при всех x и $S_n(0,p)=\int\limits_{-\pi}^{\pi} D_n(x)p(x)dx>a$.

Следующий шаг является основным.

Предложение 2. *Для любого положительного числа a найдётся неотрицательный тригонометрический многочлен q_a со свободным членом 1, такой, что*

$$\max_n |S_n(x, q_a)| > a$$

для любого x.

Прежде чем доказывать это предложение, объясним, как вывести из него основной результат.

Положим $k_j(x) = q_{j^3}(x)$, $j \in \mathbb{N}$. Это неотрицательный тригонометрический многочлен со свободным членом 1, такой, что $\max_n |S_n(x, k_j)| > j^3$ для любого x. Пусть степень этого многочлена равна N_j. Назовем *носителем* тригонометрического многочлена p любое множество S натуральных чисел, такое, что коэффициенты Фурье a_n и b_n многочлена p равны нулю при всяком $n \notin S$, $n > 0$ (свободный член не учитывается). Выберем константу $m_2 > N_1$. Тогда многочлены $k_1(x)$ и $k_2(m_2 x)$ имеют непересекающиеся носители $[1, N_1]$ и $[m_2, m_2 N_2]$. Подберем $m_3 > m_2 N_2$ и рассмотрим многочлен $k_3(m_3 x)$. Тогда три многочлена $k_1(x), k_2(m_2 x)$ и $k_3(m_3 x)$ имеют непересекающиеся носители. Далее будем поступать аналогично и построим семейство многочленов $k_j(m_j x)$, $j \in \mathbb{N}$ с непересекающимися носителями. Положим $m_1 = 1$ и, наконец, пусть

$$K(x) = \sum_{j=1}^{\infty} k_j(m_j x)/j^2.$$

Это и есть искомая функция.

Надо только проверить, что она интегрируема, понять, каков у неё ряд Фурье, и доказать, наконец, что он расходится в каждой точке.

Свободный член тригонометрического многочлена равен делённому на 2π интегралу по отрезку $[-\pi; \pi]$. Таким образом, $\frac{1}{2\pi} \int\limits_{-\pi}^{\pi} k_j(m_j x)\, dx = 1$, так что ряд $\sum\limits_{j=1}^{\infty} \int\limits_{-\pi}^{\pi} k_j(m_j x)/j^2\, dx$ сходится (его сумма равна $2\pi \sum\limits_{j=1}^{\infty} 1/j^2 = \pi^3/3$). Следовательно, функция K интегрируема. Ее ряд Фурье выглядит следующим образом: сперва идёт свободный член $\pi^2/6$, затем ненулевые члены тригонометрических многочленов $k_1(m_1 x), k_2(m_2 x)/4, \ldots$. Расходимость этого ряда в каждой точке вытекает из более сильного утверждения: $\sup_n |S_n(x, K)| = \infty$ при любом x. Зафиксируем x и покажем, что некоторая сумма $S_n(x, K)$ ряда Фурье функции K по модулю больше, например, четырёх. Найдётся такой номер l, при котором

$|S_l(m_9x, k_9/9^2)| > 9$ (предложение 2). Сумма членов ряда Фурье функции K с номерами от m_9 до m_9l в точке x совпадает с $S_l(m_9x, k_9/9^2) - 1/9^2$ и, следовательно, по модулю больше восьми. Таким образом, разность двух чисел вида $S_n(x, K)$ по модулю больше восьми, так что одно из этих чисел по модулю должно быть больше четырёх.

Пример построен, и остается только доказать предложение 2. Идею доказательства мы узнали от С. В. Конягина, который указал нам на работу Ш. В. Хеладзе [12]; мы выражаем ему благодарность за это и за ценные обсуждения.

В силу предложения 1 существует тригонометрический многочлен $p = p_{8a}$, который в каждой точке по модулю не превосходит единицы, а в нуле какая-то из его сумм Фурье больше $8a$. В силу непрерывности найдётся число $\delta > 0$ такое, что $\max\limits_n |S_n(x, p)| > 8a$ при всех $x \in (-\delta; \delta)$. Пусть l — степень многочлена p и $N > 2l$. Положим $t_1(x, N) = p(x)\cos Nx$ и $t_2(x, N) = p(x)\cos 2Nx$. Ясно, что t_1 и t_2 являются тригонометрическими многочленами без свободного члена и что они всюду не превосходят единицу по модулю. Носитель t_1 содержится в интервале $[N - l, N + l]$, а носитель t_2 — в интервале $[2N - l, 2N + l]$. Это следует из формул $\cos x \cos y = (\cos(x+y)+\cos(x-y))/2$ и $\sin x \cos y = (\sin(x+y)+\sin(x-y))/2$.

Покажем, что для всякого $x \in (-\delta, \delta)$ выполняется по меньшей мере одно из неравенств: $\max\limits_{m,n} |S_m(x, t_1) - S_n(x, t_1)| > 4a$ или $\max\limits_{m,n} |S_m(x, t_2) - S_n(x, t_2)| > 4a$. Пусть задано $x \in (-\delta; \delta)$. Найдётся $n \leqslant l$, для которого $|S_n(x, p)| > 8a$. Заметим, что для любого y одно из чисел $\cos y$ и $\cos 2y$ по модулю $\geqslant 1/2$ (это следует из формулы $\cos 2y = 2\cos^2 y - 1$). Таким образом, выполняется одно из неравенств: $|\cos Nx| \geqslant 1/2$ или $|\cos 2Nx| \geqslant 1/2$. Предположим, например, что имеет место первое из них. Сумма $S_{N+n}(x, t_1) - S_{N-n-1}(x, t_1)$ членов многочлена t_1 с номерами от $N - n$ до $N + n$ в точке x совпадает с $S_n(x, p)\cos Nx$ и потому по модулю $> 8a/2 = 4a$. Аналогично разбирается случай $|\cos 2Nx| \geqslant 1/2$.

Выберем точки $c_1, \ldots, c_s$ на отрезке $[-\pi; \pi]$ так, чтобы δ-окрестности этих точек покрывали весь отрезок. Положим $r_i(x, N) = t_1(x - c_i, N)$ и $r'_i(x, N) = t_2(x - c_i, N)$, $1 \leqslant i \leqslant s$. Тогда r_i и r'_i — тригонометрические многочлены с носителями $[N - l, N + l]$ и $[2N - l, 2N + l]$ соответственно. Для каждого $x \in [-\pi; \pi]$ найдётся такое i, $1 \leqslant i \leqslant s$, что одно из чисел $\max\limits_{m,n} |S_m(x, r_i) - S_n(x, r_i)|$ и $\max\limits_{m,n} |S_m(x, r'_i) - S_n(x, r'_i)|$ превосходит $4a$. Это вытекает из предыдущего абзаца и того, что $S_n(x, r_i) = S_n(x - c_i, t_1)$, $S_n(x, r'_i) = S_n(x - c_i, t_2)$.

Пусть

$$q(x) = \prod_{i=1}^{s}(1 + (r_i(x, N_i) + r_i'(x, N_i))/2),$$

где о натуральных числах $N_1, \dots, N_s$ нам ещё предстоит позаботиться. Тогда q — неотрицательный тригонометрический многочлен. Числа $N_1, \dots, N_s$ выберем шаг за шагом так, чтобы все многочлены $r_i(x, N_i)/2$ и $r_i'(x, N_i)/2$ «содержались бы в q» в том смысле, что коэффициенты Фурье с номерами от $N_i - l$ до $N_i + l$ у многочлена q были бы такими же, как у многочлена $r_i/2$, а коэффициенты с номерами от $2N_i - l$ до $2N_i + l$ — такими же, как у $r_i'/2$. Пусть числа $N_1, \dots, N_k$ уже выбраны так, что $r_i/2$ и $r_i'/2$, $i = 1, \dots, k$, содержатся в указанном смысле в произведении $q_k = \prod_{i=1}^{k}(1+(r_i+r_i')/2)$, причем свободный член многочлена q_k равен единице, а его носитель не пересекается с интервалом $[1, 2l]$ (на первом шаге конструкции это условие будет обеспечено, если мы выберем $N_1 > 3l$). Пусть M_k — степень многочлена q_k. Выбираем $N_{k+1} > 2M_k + 2l$. Тогда условия, которые мы накладывали на q_k, остаются в силе и для q_{k+1}. Действительно, $q_{k+1} = q_k(1 + (r_{k+1} + r_{k+1}')/2) = q_k + r_{k+1}/2 + r_{k+1}'/2 + R$, где $R = (q_k - 1)(r_{k+1} + r_{k+1}')/2$. Полиномы $q_k - 1$ и $r_{k+1} + r_{k+1}'$ не имеют свободных членов, носитель первого из них содержится в интервале $I_1 = [2l + 1, M_k]$, носитель второго — в объединении I_2 интервалов $[N_{k+1} - l, N_{k+1} + l]$ и $[2N_{k+1} - l, 2N_{k+1} + l]$, поэтому носитель многочлена R содержится в множестве $I_2 \pm I_1$ всех чисел, представимых в виде $m + n$ или $m - n$, где $m \in I_2$, $n \in I_1$. В силу нашего выбора числа N_{k+1} множество $I_2 \pm I_1$ не пересекается ни с $[1, M_k]$, ни с I_2. Таким образом, многочлены q_k, $r_{k+1}/2$, $r_{k+1}'/2$ и R имеют попарно непересекающиеся носители и потому «содержатся» в своей сумме q_{k+1}. Свободный член многочлена q_{k+1} равен единице. Тем самым конструкция завершена.

Тот факт, что многочлены $r_i/2$ содержатся в q, означает, что каждая сумма $S_m(x, r_i/2) - S_n(x, r_i/2)$ идущих подряд членов многочлена $r_i/2$ совпадает с суммой $S_m(x, q) - S_n(x, q)$ идущих подряд членов многочлена q, и аналогично для r_i'. Мы видели, что для каждой точки x некоторая сумма такого вида превосходит $2a$. Следовательно, $\max\limits_n |S_n(x, q)| > a$, так что q удовлетворяет требованиям предложения 2.

Л. С. Понтрягин и топологическая теорема двойственности

Лев Семёнович Понтрягин (1908 – 1988) внес выдающийся вклад во многие разделы математики: теорию дифференциальных уравнений, теорию оптимального управления, топологическую алгебру, геометрию, теорию групп Ли, но, прежде всего, – в топологию, с которой он начинал. Он, безусловно, принадлежит к числу великих русских математиков и величайших топологов нашего века.

В возрасте 14 лет от взорвавшегося примуса Понтрягин потерял зрение. В 1925 году он поступает в Московский университет и становится учеником П. С. Александрова. В студенческие годы Понтрягин получает выдающийся результат — обобщение закона двойственности Александера. В сборнике «Математика в СССР за 15 лет», изданном в 1932 году, Понтрягин, которому было тогда 24 года, упоминается 23 раза (больше него только М. А. Лаврентьев — 24 раза), и это показатель его выдающегося вклада в нашу науку уже на заре его творческой деятельности.

Мы расскажем здесь о топологической теореме двойственности Понтрягина. Она связана с другой теоремой двойственности Понтрягина — для локально компактных групп, о ней мы тоже упомянем. Обзор творчества Л. С. Понтрягина можно найти в [1]. Мы отсылаем читателя также к обзору самого Понтрягина «О моих работах по топологии и топологической алгебре» (см. [8]).

Известная теорема Жордана утверждает, что *простая замкнутая кривая разбивает плоскость на две области.* Напомним, что простая замкнутая кривая — это множество, гомеоморфное окружности. Теорема Жордана допускает обобщение на случай пространств любой размерности: всякое подмножество евклидова пространства $\mathbb{R}^n$, гомеоморфное сфере S^{n-1}, имеет дополнение, состоящее из двух компонент. Более того, *если замкнутые подмножества F_1 и F_2 пространства $\mathbb{R}^n$ гомеоморфны, то их дополнения имеют одинаковое число компонент.* Выделенное курсивом утверждение — это весьма частный случай теоремы двойственности Понтрягина, которая устанавливает связь между топологическими свойствами произвольного замкнутого подмножества евклидова пространства (или многомерной сферы) и его дополнения.

Формулировка теоремы Понтрягина использует понятие *группы гомологий* топологического пространства. Мы не будем давать общего определения, а ограничимся случаем открытых подмножеств евклидовых пространств.

Пусть X — открытое подмножество плоскости. Нульмерная цепь в X (или просто 0-цепь) — это конечное множество точек в X, каждой

из которых приписано целое число. Такую цепь будем записывать как формальную сумму точек из X с целыми коэффициентами. Аналогично, 1-цепь или 2-цепь в X — это совокупность ориентированных отрезков (соответственно треугольников), целиком лежащих в X, каждому из которых приписано целое число. Более формально, совокупность 1-цепей — это свободная абелева группа, порождённая множеством всех ориентированных отрезков, лежащих в X, а совокупность 2-цепей — это свободная абелева группа, порождённая множеством всех ориентированных треугольников, лежащих в X. (Напомним, что свободная абелева группа, порождённая множеством Y, — это множество формальных сумм вида $n_1y_1 + n_2y_2 + \cdots + n_sy_s$, $n_i \in \mathbb{Z}$, $y_i \in Y$.)

Граница 1-цепи — это 0-цепь, определяемая так: граница ориентированного отрезка $[a;b]$ (где a и b — точки плоскости) — это формальная разность $b - a$ («конец минус начало»), и граница суммы равна сумме границ. Аналогично, граница ориентированного треугольника abc — это 1-цепь $[a;b]+[b;c]+[c;a]$, а граница произвольной 2-цепи определяется по линейности.

1-цепь с нулевой границей называется *1-циклом*. Примером 1-цикла служит любая замкнутая ломаная, то есть 1-цепь вида $[a_1;a_2]+[a_2;a_3]+\\+\cdots+[a_{n-1};a_n]+[a_n;a_1]$. Нетрудно понять, что 1-циклы — это в точности суммы замкнутых ломаных с целыми коэффициентами.

1-цепь называется *1-границей*, если она является границей некоторой 2-цепи. Так как граница любого ориентированного треугольника является замкнутой ломаной и, следовательно, 1-циклом, то и *всякая 1-граница является 1-циклом*. Эквивалентно, *граница границы равна нулю*. В этом заключается важнейшее свойство введённого выше оператора границы.

Всякий ли 1-цикл является 1-границей? Ответ зависит от рассматриваемого открытого множества X. Например, если X — это вся плоскость или внутренность круга, то ответ положителен. Простейший пример ситуации, когда цикл не является границей, доставляет проколотая плоскость. Пусть p — точка на плоскости, X — дополнение до p. Рассмотрим замкнутую ломаную c в X, обходящую вокруг p. Например, нашей ломаной c может быть граница треугольника, содержащего p в своей внутренности. Тогда c не является 1-границей в X. Геометрически это очевидно. Для формального доказательства заметим, что каждому 1-циклу z в X можно соотнести целое число — число оборотов цикла вокруг точки p. Имея в виду дальнейшие обобщения, мы будем называть это число также *индексом зацепления* 1-цикла z и 0-цикла p. Это число можно определить, например, так. Пусть $z = [a_1;a_2]+\cdots+[a_n;a_1]$ — замкнутая ломаная в X. Проведем луч из p, не проходящий через вершины $a_1, \ldots, a_n$ ломаной z,

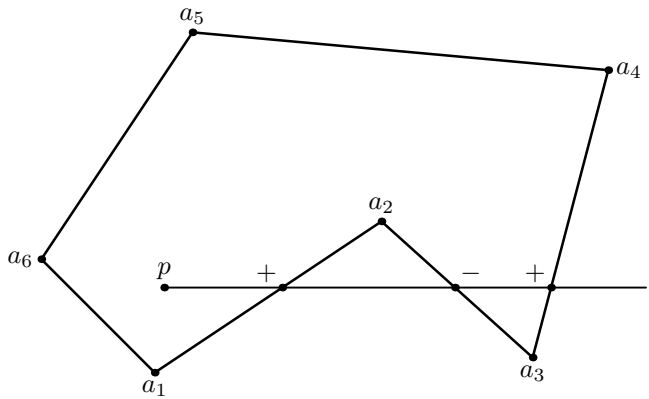


Рис. 1.

и каждой точке пересечения луча с ломаной припишем число ± 1 по следующему правилу: если звено $[a_i; a_{i+1}]$ пересекает луч справа налево, то точке пересечения приписывается плюс 1, в противном случае — минус 1. Число оборотов ломаной z вокруг p определяется как сумма плюс-минус единиц по всем точкам пересечения (см. рис. 1). Легко видеть, что это число корректно определено (оно не меняется при вращении луча) и что наше определение приводит к тому, что мы понимаем под «числом оборотов». Это же определение годится для любого 1-цикла, не обязательно являющегося замкнутой ломаной.

Очевидно, граница любого треугольника в X (то есть треугольника, не содержащего точку p) делает 0 оборотов вокруг p. Это следует из данного выше определения, использующего пересечения с лучами: луч, исходящий из точки вне треугольника и не проходящий через вершины, либо вообще не пересекается с треугольником, либо пересекается ровно с двумя сторонами, причем двум точкам пересечения приписываются противоположные знаки. В силу аддитивности числа оборотов любая 1-граница в X делает 0 оборотов вокруг p. Следовательно, ломаная c, для которой число оборотов вокруг p отлично от нуля, не является 1-границей.

Таким образом, мотивировано следующее определение. Два 1-цикла называются *гомологичными*, если их разность является границей. Группа *1-гомологий* $H_1(X)$ пространства X — это группа классов гомологичных 1-циклов. Иными словами, если Z_1 — это группа 1-циклов, а B_1 — группа 1-границ, то $H_1 = H_1(X)$ — это факторгруппа Z_1/B_1. Наше определение группы $H_1(X)$ не было топологически инвариантным: не сразу видно, почему гомеоморфные открытые множества на плоскости имеют изоморфные группы гомологий. На самом деле можно так видоизменить

определение, что топологическая инвариантность групп гомологий станет очевидной: при определении цепей надо рассматривать не вложенные в X отрезки и треугольники, а произвольные непрерывные отображения фиксированного отрезка или треугольника в X. Группы цепей при этом увеличиваются, но можно доказать, что группы гомологий остаются теми же.

Аналогичные определения можно дать для любых размерностей. Пусть X — открытое множество в $\mathbb{R}^n$. Для любого целого k определим группу C_k всех k-цепей пространства X как свободную абелеву группу, порождённую всеми ориентированными k-мерными симплексами, лежащими в X (при $k < 0$ и $k > n$ группа C_k сводится к нулю). Как и раньше, определяются граничный оператор $\partial : C_k \to C_{k-1}$, группы k-циклов Z_k, группы k-границ B_k и группы гомологий $H_k = Z_k/B_k$. Это определение допускает различные модификации. Например, вместо вложенных симплексов можно рассматривать *сингулярные симплексы*, то есть непрерывные отображения фиксированного симплекса в заданное пространство (этот подход приводит к топологически инвариантной теории), непрерывные отображения можно заменить на гладкие, симплексы — на произвольные выпуклые многогранники. Группы цепей при этом меняются, но группы гомологий остаются теми же.

Теперь мы можем сформулировать топологическую теорему двойственности Понтрягина.

Топологическая теорема двойственности Понтрягина (первая формулировка). *Пусть F_1 и F_2 — гомеоморфные замкнутые множества, лежащие в $\mathbb{R}^n$. Тогда их дополнения имеют изоморфные группы гомологий; иными словами, группы $H_k(\mathbb{R}^n \setminus F_1)$ и $H_k(\mathbb{R}^n \setminus F_2)$ изоморфны при любом k.*

Этот результат обобщает теорему американского математика Дж. Александера (1888–1971), доказанную им в 1922 году. Александер вместо произвольных замкнутых множеств в $\mathbb{R}^n$ рассматривал множества, гомеоморфные полиэдрам, а вместо изоморфизма групп у него фигурировало равенство чисел Бетти (это некоторые численные характеристики групп гомологий).

Таким образом, с каждым компактом, расположенным в $\mathbb{R}^n$, можно связать новые топологические инварианты: группы гомологий дополнения. Эти группы изоморфны так называемым *группам когомологий* рассматриваемого компакта. Группы когомологий были введены в рассмотрение в 30-е годы независимо А. Н. Колмогоровым и Дж. Александером.

Эти группы можно определить различными способами, один из них мы обсудим ниже.

Всякая 0-цепь является 0-циклом. Пусть X — открытое множество евклидова пространства. 0-цепь $\sum_{x\in X} n_x x$ является 0-границей тогда и только тогда, когда для всякой связной компоненты V множества X сумма коэффициентов $\sum_{x\in V} n_x$ по всем точкам из V равна нулю. Назовем 0-цепь $\sum_{x\in X} n_x x$ *приведённой*, если $\sum_{x\in X} n_x = 0$. Пусть $\tilde{Z}_0$ — группа приведённых 0-цепей. *Приведённая нульмерная группа гомологий* $\tilde{H}_0(X)$ — это факторгруппа $\tilde{Z}_0/B_0$ группы приведённых 0-циклов по подгруппе 0-границ. Если X имеет n компонент, то $H_0(X)$ — свободная абелева группа с n образующими, а $\tilde{H}_0(X)$ — свободная абелева группа с $n-1$ образующими.

При $k \neq 0$ полагаем $\tilde{H}_k = H_k$.

Дадим вторую формулировку теоремы сначала для «знатоков»; разъяснения для остальных мы дадим позже.

Топологическая теорема двойственности Понтрягина (вторая формулировка). *Пусть F — собственное замкнутое подмножество в $\mathbb{R}^n$, $G = \mathbb{R}^n \setminus F$ — его дополнение. Для любого целого k приведённая группа гомологий $\tilde{H}_k(G)$ изоморфна группе когомологий с компактными носителями $H_c^{n-k-1}(F)$.*

Применяя эту теорему при $k = 0$, получаем такое следствие: если F — компакт, лежащий в $\mathbb{R}^n$, то число компонент множества $\mathbb{R}^n \setminus F$ на единицу больше ранга свободной абелевой группы $H^{n-1}(F)$. В частности, если F гомеоморфно $(n-1)$-мерной сфере, то группа когомологий $H^{n-1}(F)$ изоморфна группе $\mathbb{Z}$ целых чисел, так что F разбивает $\mathbb{R}^n$ на две области. Отметим также, что компакт F не разбивает $\mathbb{R}^n$ (то есть имеет связное дополнение) тогда и только тогда, когда группа $H^{n-1}(F)$ нулевая. Можно показать, что последнее условие равносильно такому: пространство $(S^{n-1})^F$ всех непрерывных отображений из F в $(n-1)$-мерную сферу S^{n-1}, наделённое естественной метрикой, связно. Таким образом, компакт F разбивает $\mathbb{R}^n$ тогда и только тогда, когда некоторое отображение $f : F \to S^{n-1}$ нельзя соединить с постоянным отображением $g : F \to S^{n-1}$ непрерывным путем в пространстве $(S^{n-1})^F$ (П. С. Александров). Если F разбивает $\mathbb{R}^n$, то в качестве f можно взять центральную проекцию на расположенную в $\mathbb{R}^n$ сферу с центром p, где p — произвольная точка, принадлежащая какой-либо ограниченной компоненте множества $\mathbb{R}^n \setminus F$.

Короткий путь к доказательству топологической теоремы двойственности лежит через алгебраические понятия комплекса и точной последовательности. *Цепной комплекс* — это последовательность абелевых групп

C_k и гомоморфизмов $d_k : C_k \to C_{k-1}$, таких, что $d_{k-1}d_k = 0$ при всех k. Пусть $Z_k = \{x \in C_k : d_k x = 0\}$ — *группа циклов* и $B_k = d_{k+1}(C_{k+1})$ — *группа границ*. Факторгруппа $H_k = Z_k/B_k$ — это *группа гомологий степени* k рассматриваемого комплекса. Например, гомологии топологического пространства — это гомологии некоторого комплекса, связанного с данным пространством. Аналогично определяются коцепные комплексы. Это по существу то же самое, что и цепные комплексы, только нумерация групп меняется на обратную: гомоморфизмы δ_k действуют из C_k в C_{k+1} и удовлетворяют соотношению $\delta_{k+1}\delta_k = 0$. В этом случае говорят о коциклах, кограницах и когомологиях.

Естественным образом определяются понятия подкомплекса и факторкомплекса. Пусть K — цепной комплекс, K' — его подкомплекс, K'' — соответствующий факторкомплекс. Тогда гомологии комплексов K, K' и K'' связаны последовательностью гомоморфизмов

$$\dots \to H_{n+1}(K'') \to H_n(K') \to H_n(K) \to H_n(K'') \to H_{n-1}(K') \to \dots,$$

причем эта последовательность *точна*, то есть ядро каждого гомоморфизма совпадает с образом предыдущего. (В одном учебнике по поводу этого утверждения сказано: «Читатель должен один и только один раз в своей жизни проследить все детали до конца».) В случае коцепных комплексов соответствующая *длинная точная последовательность когомологий* имеет вид

$$\dots \to H^{n-1}(K'') \to H^n(K') \to H^n(K) \to H^n(K'') \to H^{n+1}(K') \to \dots.$$

Пусть теперь X — локально компактное пространство (например, открытое или замкнутое подмножество евклидова пространства). Определим, следуя [6], *группы когомологий с компактными носителями* $H_c^k(X)$ пространства X. Фиксируем $k \geqslant 0$. Пусть $\Phi^k(X)$ — группа всех отображений из X^{k+1} в $\mathbb{Z}$ (без каких-либо условий непрерывности). *Носитель* отображения $\varphi \in \Phi^k(X)$ — это замкнутое множество всех $x \in X$, таких, что для всякой окрестности U точки x найдутся $x_0, \dots, x_k \in U$ с $\varphi(x_0, \dots, x_k) \neq 0$. Пусть $\Phi_c^k(X)$ — группа всех $\varphi \in \Phi^k(X)$ с компактным носителем, $\Phi_0^k(X)$ — группа всех $\varphi \in \Phi^k(X)$ с пустым носителем. Факторгруппа $C_c^k(X) = \Phi_c^k(X)/\Phi_0^k(X)$ называется *группой k-мерных коцепей с компактными носителями*.

Кограничный гомоморфизм $\delta_k : \Phi^k(X) \to \Phi^{k+1}(X)$ определяется формулой

$$(\delta_k\varphi)(x_0, \dots, x_{k+1}) = \sum_{i=0}^{k+1}(-1)^i\varphi(x_0, \dots, \hat{x}_i, \dots, x_{k+1}),$$

где крышка $\hat{}$ над символом означает, что его следует пропустить. Например, если $\varphi \in \Phi^1(X)$, то $\delta_1\varphi$ определяется формулой $(\delta_1\varphi)(a,b,c) = \varphi(b,c) - \varphi(a,c) + \varphi(a,b)$. Гомоморфизм δ_k не увеличивает носители и потому отображает $\Phi_c^k(X)$ в $\Phi_c^{k+1}(X)$ и $\Phi_0^k(X)$ в $\Phi_0^{k+1}(X)$. При переходе к факторгруппам возникает гомоморфизм групп коцепей $C_c^k(X) \to C_c^{k+1}(X)$, который мы обозначаем по-прежнему через δ_k.

Гомоморфизм $\delta_{k+1}\delta_k$ равен нулю: в формулу для $\delta_{k+1}\delta_k\varphi$ каждое слагаемое вида $\varphi(x_0, \ldots, \hat{x}_i, \ldots, \hat{x}_j, \ldots, x_{k+2})$ входит два раза с противоположными знаками. Следовательно, группы коцепей $C_c^k(X)$ вместе с кограничными гомоморфизмами δ_k образуют коцепной комплекс $C_c^*(X)$. Группы когомологий этого комплекса и называются группами когомологий с компактными носителями пространства X.

Пусть F — замкнутое подмножество в X, $G = X \setminus F$. Ограничение коцепей определяет эпиморфизм комплексов $C_c^*(X) \to C_c^*(F)$, ядро которого в некотором смысле эквивалентно комплексу $C_c^*(G)$ и, в частности, имеет те же когомологии [6]. Возникающая длинная точная последовательность когомологий имеет вид

$$\cdots \to H_c^{n-1}(F) \to H_c^n(G) \to H_c^n(X) \to H_c^n(F) \to H_c^{n+1}(G) \to \ldots.$$

С помощью этой последовательности можно вычислить когомологии с компактными носителями евклидова пространства $\mathbb{R}^n$. Оказывается, что группа $H_c^k(\mathbb{R}^n)$ нулевая при $k \neq n$ и изоморфна $\mathbb{Z}$ при $k = n$. Это просто, если $n = 0$, а общий случай выводится отсюда так. Пусть X — замкнутое полупространство в $\mathbb{R}^n$, F — граничная гиперплоскость. Из того, что одноточечная компактификация пространства X гомеоморфна шару, выводится, что $H_c^k(X) = 0$ при всех k. Точная последовательность когомологий пары (X, F) содержит куски вида

$$0 \to H_c^k(F) \to H_c^{k+1}(X \setminus F) \to 0.$$

Так как F гомеоморфно $\mathbb{R}^{n-1}$, а $X \setminus F$ гомеоморфно $\mathbb{R}^n$, мы видим, что $H_c^k(\mathbb{R}^{n-1})$ изоморфно $H_c^{k+1}(\mathbb{R}^n)$, так что всё сводится к случаю точки.

Пусть теперь F — замкнутое подмножество в $\mathbb{R}^n$, $G = \mathbb{R}^n \setminus F$ — его дополнение. Точная когомологическая последовательность пары $(\mathbb{R}^n, F)$ содержит куски вида

$$H_c^k(\mathbb{R}^n) \to H_c^k(F) \to H_c^{k+1}(G) \to H_c^{k+1}(\mathbb{R}^n).$$

При $k \neq n, n-1$ по краям стоят нули, откуда следует изоморфизм $H_c^k(F) \cong H_c^{k+1}(G)$. Но для любого ориентируемого топологического n-мерного многообразия M (в частности, для открытого множества в $\mathbb{R}^n$) группа когомологий $H_c^k(M)$ изоморфна группе гомологий $H_{n-k}(M)$ (*изоморфизм*

двойственности Пуанкаре [6, теорема 11.2]). Здесь имеются в виду «гомологии с компактными носителями», которые для многообразий совпадают с сингулярными. Таким образом, группы $H_c^{k+1}(G)$ и $H_{n-k-1}(G)$ изоморфны, и мы приходим к требуемому изоморфизму $H_c^k(F) \cong H_{n-k-1}(G)$. При $k = n-1$ это рассуждение нуждается в некоторой модификации: из точной последовательности

$$0 \to H_c^{n-1}(F) \to H_c^n(G) \to H_c^n(\mathbb{R}^n) \cong \mathbb{Z} \to H_c^n(F) = 0$$

и изоморфизма двойственности Пуанкаре $H_c^n(G) \cong H_0(G)$ следует изоморфизм $H_c^{n-1}(F) \cong \tilde{H}_0(G)$.

Таким образом, мы вывели двойственность Понтрягина из двойственности Пуанкаре. Что касается изоморфизма Пуанкаре, то наиболее естественное доказательство использует пучки абелевых групп и их когомологии. Эти понятия, введённые французской математической школой в 50-е годы, преобразили алгебраическую геометрию и многомерный комплексный анализ и играют центральную роль в современном построении этих разделов математики. Именно теория пучков, включённая в контекст общего понятия *абелевой категории*, даёт естественное объяснение понятия когомологии топологического пространства. О теории пучков можно узнать из [3, 2]. Теорему об изоморфизме $H_c^k(M) \cong H_{n-k}(M)$, которая теперь называется теоремой Пуанкаре, сам Пуанкаре формулировал иначе: ведь когомологии были изобретены спустя два десятилетия после его смерти. В формулировке Пуанкаре фигурировало равенство чисел Бетти. Доказательство теоремы двойственности Пуанкаре, основанное на пучках, изложено в [2] и в добавлениях Е. Г. Скляренко к русским переводам книг [6] и [2].

Наш «алгебраизированный» подход к двойственности Понтрягина затушёвывает геометрическую сущность этой двойственности. Между тем изоморфизм между $\tilde{H}_k(\mathbb{R}^n \setminus F)$ и $H_c^{n-k-1}(F)$ имеет геометрическое истолкование и может быть выражен в терминах «зацепления циклов». С примерами индексов зацепления мы уже встречались, когда обсуждали понятие числа оборотов 1-цикла вокруг точки на плоскости. В общем случае (целочисленный) индекс зацепления между непересекающимися k- и $(n-k-1)$-циклами в $\mathbb{R}^n$ определяется как индекс пересечения одного из циклов с цепью, границей которой является другой цикл. При этом индекс пересечения между «находящимися в общем положении» пересекающимися k- и $(n-k)$-мерными симплексами полагается равным ± 1 в зависимости от ориентации, а в случае произвольных цепей надо плюс-минус единицы просуммировать.

Если F — компактный полиэдр (= объединение конечного числа симплексов) в $\mathbb{R}^n$, то при некоторых дополнительных предположениях о F группу когомологий $H^{n-k-1}(F)$ можно отождествить с группой гомоморфизмов из $H_{n-k-1}(F)$ в $\mathbb{Z}$. Изоморфизм между $\tilde{H}_k(\mathbb{R}^n \setminus F)$ и $H_c^{n-k-1}(F)$ допускает тогда следующее описание: каждому циклу $c \in \tilde{H}_k(\mathbb{R}^n \setminus F)$ сопоставляется гомоморфизм $H_{n-k-1}(F) \to \mathbb{Z}$, который каждому $(n-k-1)$-циклу на F соотносит его индекс зацепления с c. В общем случае геометрическая интерпретация топологической теоремы двойственности Понтрягина связана с совершенно другой «теоремой двойственности Понтрягина» — теоремой двойственности для локально компактных абелевых групп.

Обозначим через $\mathbb{T}$ мультипликативную группу комплексных чисел, равных по модулю единице, или изоморфную группу $\mathbb{R}/\mathbb{Z}$. Это — компактная топологическая группа. Для произвольной локально компактной абелевой топологической группы G обозначим через G^* группу всех непрерывных гомоморфизмов из G в $\mathbb{T}$, наделённую топологией равномерной сходимости на компактных множествах. Тогда G^* — локально компактная абелева группа, называемая двойственной для G. Имеется естественный гомоморфизм из G в G^{**}. Теорема двойственности Понтрягина для локально компактных групп утверждает, что: (1) этот гомоморфизм является изоморфизмом топологических групп; (2) переход от G к G^* устанавливает взаимнооднозначное соответствие между компактными и дискретными абелевыми группами: группа G^* компактна тогда и только тогда, когда G дискретна (и наоборот). Например, двойственными являются группа целых чисел $\mathbb{Z}$ и окружность $\mathbb{T}$, а группа $\mathbb{R}$ действительных чисел двойственна сама себе. Понятие двойственной по Понтрягину группы позволяет с единой точки зрения рассматривать ряды Фурье и преобразование Фурье на прямой: в общем случае «преобразование Фурье» отображает меры или функции на локально компактной группе G в функции на G^*. Элементарное изложение теоремы двойственности для локально компактных групп можно найти в [7].

Пусть теперь F — компактное подмножество в $\mathbb{R}^n$. Тогда группу, двойственную к дискретной группе когомологий $H^k(F)$, можно интерпретировать как «группу гомологий с коэффициентами в $\mathbb{T}$» множества F. Обозначим эту компактную группу через $H_k(F, \mathbb{T})$. Элементы группы $H_k(F, \mathbb{T})$ представляются «циклами с компактными коэффициентами», расположенными в сколь угодно малой окрестности компакта F. Не вдаваясь в детали, укажем лишь, что можно естественным образом определить «индекс зацепления» между k-циклами в $\mathbb{R}^n \setminus F$ и элементами группы $H_{n-k-1}(F, \mathbb{T})$. Этот индекс зацепления принимает значения в $\mathbb{T}$.

При определении гомологий открытых множеств евклидовых пространств мы пользовались выше цепями с целочисленными коэффициентами. Аналогично можно определить цепи и гомологии с коэффициентами в произвольной абелевой группе A, а также когомологии с коэффициентами в A — для этого при определении коцепей надо рассматривать функции со значениями в A.

Топологическая теорема двойственности Понтрягина (третья формулировка). *Пусть F — компакт, лежащий в $\mathbb{R}^n$, $G = \mathbb{R}^n \setminus F$ — его дополнение. Для любой абелевой группы A и целого числа k индекс зацепления устанавливает двойственность между дискретной группой $\check{H}_k(G, A)$ и компактной группой $H_{n-k-1}(F, A^*) = (H^{n-k-1}(F, A))^*$.*

Мы не будем более подробно объяснять смысл этой теоремы или доказывать её, а ограничимся некоторыми примерами.

Пример 1. Пусть Ω — открытое множество на плоскости, c — 1-цикл в Ω. Мы видели, что следующее условие необходимо для того, чтобы c было 1-границей в Ω: если p — точка плоскости, не лежащая в X, то c не обходит вокруг p (то есть делает 0 оборотов вокруг p). Это условие также и достаточно. Этот факт допускает элементарное доказательство, но может рассматриваться и как частный случай теоремы двойственности. Он важен в теории функций комплексного переменного. Пусть $\Omega \subset \mathbb{C}$ и f — функция, голоморфная в Ω. Если 1-цикл c в Ω таков, что для любой точки $p \in \mathbb{C} \setminus \Omega$ цикл c не обходит вокруг p, то c является 1-границей в Ω и потому интеграл $\int\limits_c f(z)\,dz$ равен нулю (так как он равен сумме интегралов по границам треугольников, целиком расположенных в Ω).

Пример 2. Пусть F — компакт в $\mathbb{R}^n$, содержащий конечное число s связных компонент $F_1, \ldots, F_s$, и пусть $G = \mathbb{R}^n \setminus F$ — его дополнение. Тогда группа гомологий $H_{n-1}(G)$ — свободная абелева группа с s образующими. Базис в $H_{n-1}(G)$ образуют $(n-1)$-циклы $c_1, \ldots, c_s$, такие, что c_i обходит 1 раз вокруг любой точки из F_i и обходит 0 раз вокруг любой точки из F_j при $j \neq i$.

Пример 3. Пусть F — узел в $\mathbb{R}^3$, то есть простая замкнутая кривая. Тогда 1-цикл в $\mathbb{R}^3 \setminus F$ гомологичен нулю тогда и только тогда, когда он имеет нулевой индекс зацепления с F. Отметим, что при этом *фундаментальная группа* пространства $\mathbb{R}^3 \setminus F$ может быть устроена достаточно сложно. Элементами фундаментальной группы являются, грубо говоря, классы непрерывных замкнутых кривых, причем две кривые принадлежат одному классу, если одну из них можно непрерывно продеформировать в другую.

На рисунке 2 (за который мы благодарим А. Б. Сосинского) изображен узел F, 1-цикл β в $G = \mathbb{R}^3 \setminus F$, имеющий нулевой индекс зацепления с F, и 1-цикл α, имеющий единичный индекс зацепления с F. Цикл β нельзя «расцепить» с F, то есть стянуть в точку, оставаясь в G. В то же время по теореме двойственности цикл β гомологичен нулю в G. На рисунке показана расположенная в G поверхность S. Разбив её на треугольники, мы получим 2-цепь в G с границей β. Группа $H_1(G) \cong H^1(F) \cong \mathbb{Z}$ порождается 1-циклом α.

Пример 4. Пусть F — подмножество в $\mathbb{R}^3$, гомеоморфное двумерной сфере. Тогда группа гомологий $H_1(\mathbb{R}^3 \setminus F)$ равна нулю. При этом фундаментальная группа пространства $\mathbb{R}^3 \setminus F$ может быть отлична от нуля: в качестве F можно взять так называемую «рогатую сферу» Александера (из сферы вытягивается длинный рог и на нем завязывается бесконечное множество узелков). На плоскости подобное невозможно: если F — простая замкнутая кривая на $\mathbb{R}^2$, то существует гомеоморфизм плоскости на себя, переводящий F в обычную окружность (теорема Шенфлиса).

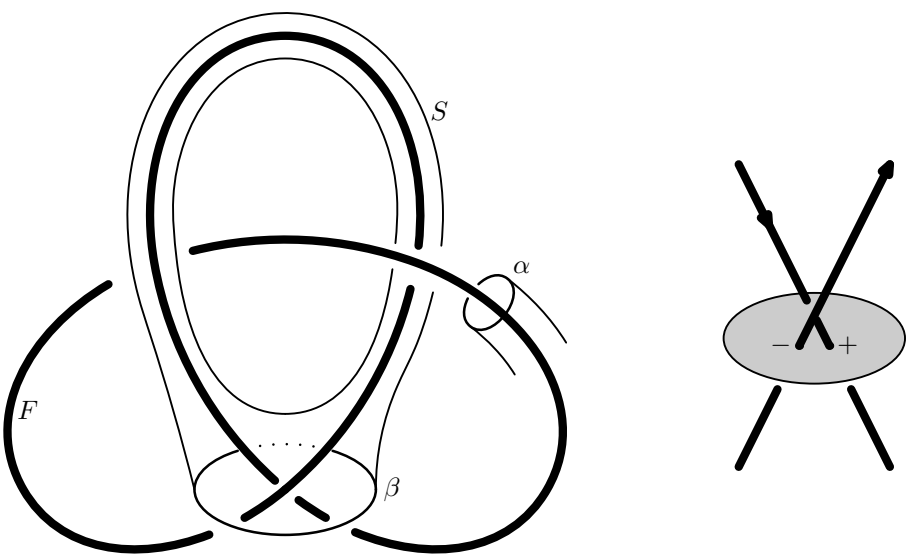


Рис. 2.

Список литературы

[1] *Аносов Д. В., Гамкрелидзе Р. В., Мищенко Е. Ф., Постников М. М.* О математических трудах Л. С. Понтрягина // Л. С. Понтрягин. Избранные научные труды. Том 1. Топология. Топологическая алгебра. М: Наука, 1988. С. 10–26.

[2] *Бредон Г.* Теория пучков. М.: Мир, 1988.

[3] *Годеман Р.* Алгебраическая топология и теория пучков. М.: Изд-во иностр. лит-ры, 1961.

[4] *Колмогоров А. Н.* Избранные труды. Математика и механика. М.: Наука, 1985.

[5] *Колмогоров А. Н., Фомин С. В.* Элементы теории функций и функционального анализа. М.: Наука, 1989.

[6] *Масси У.* Теория гомологий и когомологий. М.: Мир, 1981.

[7] *Моррис С.* Двойственность Понтрягина и строение локально компактных абелевых групп. М.: Мир, 1980.

[8] *Понтрягин Л. С.* Избранные научные труды. Том 1. Топология. Топологическая алгебра. М.: Наука, 1988.

[9] УМН. 1988. Т. 43, вып. 6.

[10] *Фоменко А. Т., Фукс Д. Б.* Курс гомотопической топологии. М.: Наука, 1989.

[11] Колмогоров в воспоминаниях / под ред. А. Н. Ширяева. М.: Наука, 1993.

[12] *Хеладзе Ш. В.* О расходимости всюду рядов Фурье функций из класса $L\phi(L)$. Труды Тбил. Мат. Ин-та АН ГССР. 1989. Т. 89. С. 51–59.

Размышления о первых московских математических олимпиадах

В. М. Тихомиров

Как всё это начиналось?

Как известно, первая математическая олимпиада была проведена в Ленинграде в 1934 году. Её организовали и провели Борис Николаевич Делоне и Григорий Михайлович Фихтенгольц. Мы надеемся, что наши санкт-петербургские коллеги напишут историю своих начальных олимпиад для одного из наших следующих номеров. А здесь будет рассказано о первых математических олимпиадах[1)], состоявшихся в Москве.

Помимо личных воспоминаний участников первых олимпиад, я буду пользоваться одним «реликтовым» литературным источником — небольшой брошюрой Р. Н. Бончковского «Московские математические олимпиады 1935 и 1936 гг.»

Редакция второй серии «Математического просвещения» сочла своим долгом поместить портрет Р. Н. Бончковского и дать краткую справку о нем на первых страницах первого номера журнала (см. «Математическое просвещение». Математика, ее преподавание и история. М.: ГИТТЛ, 1957).

И мы, воспринимая своё издание, как продолжение и первой, и второй серий «Математического просвещения», хотим воспользоваться возможностью напомнить читателям нашего сборника о судьбе редактора и издателя первой серии.

Ростислав Николаевич Бончковский (1905 – 1942) — способный математик и талантливый популяризатор — был инициатором создания первой серии довоенных сборников «Математического просвещения». Эти сборники начали издаваться в 1934 году (всего вышло 13 сборников), и Ростислав Николаевич был их бессменным редактором. С началом войны издание сборников прекратилось. Ростислав Николаевич погиб во время Сталинградской битвы.

[1)] Далее для краткости мы будем всюду говорить «олимпиада», «олимпийский» и т. п., имея в виду исключительно *математические* олимпиады.

Приведем несколько цитат из брошюры Бончковского. Они интересны не только как свидетельство о зарождении олимпийского движения в Москве, но и как исторический документ. По ходу дела будем кое-что комментировать.

«Первая московская математическая олимпиада 1935 г., — пишет Р. Н. Бончковский, — была организована по инициативе Московского математического общества Наркомпросом, Московским государственным университетом и школьным отделом ГОРОНО[2)].»

«В организационный комитет вошли многие видные московские профессора и педагоги: проф. П. С. Александров (председатель комитета), директор Математического института проф. А. Н. Колмогоров, директор Московского университета проф. А. С. Бутягин, профессора Л. Г. Шнирельман, С. Л. Соболев, Л. А. Люстерник, Н. А. Глаголев, С. А. Яновская, Л. А. Тумаркин, А. Г. Курош, А. Р. Эйгес, Н. Ф. Четверухин, Е. С. Березанская. . .»

И здесь не обойтись без напоминаний. Возглавляет список Павел Сергеевич Александров (1896 – 1982) — крупнейший тополог, глава московской топологической школы, президент Московского математического общества. Андрей Николаевич Колмогоров (1903 – 1987) — один из крупнейших учёных нашего века; в те годы при Московском университете существовал Математический институт, и Колмогоров был его директором с 1932 года. Лев Генрихович Шнирельман, Сергей Львович Соболев, Лазарь Аронович Люстерник, Нил Александрович Глаголев, Софья Александровна Яновская, Лев Абрамович Тумаркин (в ту пору декан мехмата), Александр Геннадиевич Курош — выдающиеся профессора механико-математического факультета Московского университета. Николай Федорович Четверухин — известный геометр, профессор Педагогического института. Александр Романович Эйгес — школьный учитель (гимназический учитель П. С. Александрова), Елизавета Савельевна Березанская — известный педагог, автор учебников для школ. Сейчас никого из них уже не осталось.

Продолжим цитату. «В конце февраля организационный комитет распространил печатное обращение к школьникам и списки задач, предназначенных для подготовки к состязаниям. ⟨· · ·⟩ На состязания первого тура (происходившие 30 марта) пришло 314 человек.

По мысли организационного комитета на первом туре должен был произойти отсев лиц, имеющих явно недостаточную подготовку, поэто-

[2)] Справка: после 1917 года правительство состояло не из министерств, а Народных комиссариатов; Народный комиссариат просвещения сокращённо — Наркомпрос. ГОРОНО — городской отдел народного образования.

му задачи первого тура по своему характеру были близки к школьным задачам. Из 314 человек лишь 131 успешно выполнили работу и были допущены к участию во втором туре.

Между первым и вторым туром происходила усиленная подготовка к решающим состязаниям; необходимую помощь участники олимпиады получили на консультациях, происходивших в университете в определённые дни и часы. Кроме того, для участников олимпиады были прочитаны лекции, на которых они имели возможность познакомиться с основными идеями современной математики. Таких лекций было прочитано пять: проф. Александровым — «Бесконечность в математике», проф. Колмогоровым — «Симметрия и группы», проф. Курошем — «Об алгебраических операциях», проф. Глаголевым — «Логика и формы геометрии» и проф. Яновской «Метод полной индукции». Наконец, комитет рекомендовал участникам олимпиады посещать собрания Школьного математического кружка при Академии наук.$\langle\cdots\rangle$ Благодаря всей этой совокупности мероприятий олимпиада потеряла черты чисто спортивного состязания и приобрела большое образовательное и воспитательное значение.

На второй тур олимпиады, происходивший 6 июня явилось 120 человек. Из них 52 успешно выполнили задания.»

Прервем на короткое время это повествование. Мы видим, как в одночасье были сформированы принципы, которые потом сохранялись на протяжении многих десятилетий, а важнейшие, упрочившись, дошли до наших дней. Большое впечатление производит тематика докладов. Она охватывает широкий спектр вопросов математики (теория множеств, алгебра, геометрия и логика), которые, с одной стороны, близки и понятны школьникам, а с другой стороны, позволяют осветить проблемы современной математики. Оставалось сделать последний шаг — изменить характер школьного кружка, это и произошло чуть позже, ниже мы расскажем об этом.

А вот каков был состав участников первой олимпиады. «В олимпиаде приняло участие 314 человек, в том числе 227 школьников, 65 рабфаковцев; остальные — учащиеся курсов подготовки в вуз, школ взрослых и т.д. Средний возраст — 18,2 лет. Два наиболее юных участника имели по 14 лет; наиболее пожилой — 29 лет. Основная масса имела 16 – 20 лет. Мальчики составляли подавляюще большинство; девочек было лишь 69 человек.»

Всем им предстояли большие испытания в жизни, прежде всего — 1937 год и Война. Что стало с этими мальчиками и девочками? Как повлияло на их дальнейшую жизнь участие в кружках и олимпиадах? Мы расскажем и об этом, но сначала чуть пофилософствуем.

Зачем нужны олимпиады?

Отвлечемся немного и зададимся вопросами: а зачем всё это? кому всё это нужно? Не будем делать вид, что мы знаем ответ, поразмышляем ...

Послушаем сначала, что писал на эту тему председатель комитета первой олимпиады П. С. Александров в предисловии к брошюре Бончковского:

«В области математики СССР стоит на одном из первых мест $\langle\cdots\rangle$ Это первоклассное мировое положение советской математической науки является одним из завоеваний Октябрьской социалистической революции, $\langle\cdots\rangle$ русская математика в условиях царского режима не могла подняться до высоты одного из руководящих факторов мировой науки.»

Затронутый здесь вопрос достаточно интересен, так что стоит уделить ему некоторое внимание. Высказанная Александровым мысль (о том, что «не могла») имеет право на существование наряду со своим отрицанием.

Несомненным фактом является феномен советской математической школы, которая за пятнадцать лет встала на самом деле на первое место (а не на одно из первых мест). И причин тому было несколько. Среди них есть политические. Надо напомнить о том, что в те годы фашизм разгромил и фактически уничтожил немецкую математическую школу. Верно и то, что революция открыла доступ к образованию для широких слоёв общества и наступила пора всеобщей грамотности и тяги к культуре. Были причины как бы случайные, например, смена поколений во французской математической школе. Американская математика тогда только набирала обороты (в послевоенные годы американцы постепенно превзошли нас, а французы были как бы на равных; а сегодня можно говорить о том, что математика вообще вышла из национальных рамок). Но одной из самых существенных причин внезапного взлета советской математики (в противоположность тому, что пишет Александров) были глубокие корни русской культуры, сформировавшейся в условиях пресловутого «царского режима». Почти все учёные, составившие славу советской математики к 1935 году, получили прекрасное воспитание в интеллигентной среде и образование в первоклассных дореволюционных гимназиях. Затем они учились в университетах — Московском, Петроградском, Киевском, Казанском, Новороссийском (где вели преподавание Ляпунов, Марков, Стеклов, Гюнтер, Егоров, Лузин, Граве, Шатуновский и другие), находясь в окружении самого передового слоя русской интеллигенции. И наиболее красноречивые свидетельства ска-

занному — воспоминания самого Павла Сергеевича, написанные свыше сорока лет спустя; из них видно, как повезло ему родиться «в условиях царского режима». Но не будем больше обсуждать эту интересную тему.

Александров далее пишет:

«Основная забота о будущем советской науки, требует, чтобы ни одно математическое дарование $\langle\cdots\rangle$ не затерялось зря. Каждому из наших подрастающих талантов обеспечено полное внимание, полная и всесторонняя помощь и поддержка со стороны советского государства и всего социалистического общества нашей страны.» И далее: «Одной из наиболее действенных форм нашей помощи самым молодым дарованиям является организация олимпиады, т. е. широкого состязания, широкого социалистического соревнования всех наших школьников, одарённых математически и интересующихся математикой. Это состязание должно заставить лучших из них почувствовать себя уже настоящими математиками, будущими учёными. Оно должно укрепить их веру в себя, зажечь их научный энтузиазм и в то же время заставить их почувствовать, что лишь длинный путь упорной работы приведёт их к цели, к участию в качестве квалифицированных математиков, а иногда и больших самостоятельных учёных в той громадной стройке социализма, которая развернулась в нашей стране.»

И снова перед нами несколько интересных тем. Тому, в какой мере действительно «была обеспечена всесторонняя помощь и поддержка каждому из наших подрастающих талантов», мы уделим внимание, когда коснемся судьбы участников и победителей олимпиад, а сейчас несколько слов о взаимодействии личности и государства. О пользе соревнования. И об олимпиадах.

... Около сорока лет тому назад, во время моих «блужданий среди цветущей черемухи по Заонежью» с моим учителем Андреем Николаевичем Колмогоровым (цитата принадлежит ему) как-то зашла речь о том, на каких принципах должно базироваться разумное государство. И Андрей Николаевич произнес слова, к которым я не был тогда подготовлен: «*Должен соблюдаться принцип Свободы*». (И нам обоим было ясно, что такое время при нашей жизни не наступит).

Сейчас я склонен толковать слова Колмогорова так: *оба* должны быть свободны — и личность, и государство — но и тот, и другое должны иметь определённые обязательства друг перед другом. Личность должна уважать Законы (отдельный вопрос, как создать разумную законодательную систему), государство должно обеспечивать личности возможность существовать. Государство не должно препятствовать, скажем,

обогащению личности, если при этом законы не нарушаются; оно может, разумеется, особо поощрять материально тех, кто ему служит (чиновничество, полицию, армию и т. п.). Но при этом личность должна иметь право «Никому/Отчёта не давать $\langle\cdots\rangle$/для власти, для ливреи/Не гнуть ни совести, ни помыслов, ни шеи» и т. д., и эти права должны быть гарантированы разумно устроенным государством. (Но и личность, «не гнущая шеи», не должна особенно сетовать на то, что государство её не слишком вознаграждает.)

Так что *служение государству* должно быть добровольным, и оно не может быть целью просвещения, в частности, целью проведения олимпиад. Одной из задач первого и непременной целью второго является *возжигание огня* (в душах) во имя процветания культуры и всего человечества в целом.

Что же касается спортивной компоненты олимпиад (того, что было названо «социалистическим соревнованием»), то здесь необходимы оговорки.

Довольно спорной является идея «рейтингования» (очень распространённая ныне), при которой человек получает талон с каким-то номером, и отоваривается согласно этому талону. И вообще человек в разумном государстве не должен быть поставлен в положение, когда ему постоянно приходится утверждать себя в конкурентной борьбе. Основные права и свободы, в частности, на образование, получение информации, на свободомыслие и т. п. должны гарантироваться без конкурсов. И потому «своим успехам на олимпиаде естественно радоваться и даже гордиться ими. Неудачи же на олимпиадах не должны чрезмерно огорчать»,— так писал Андрей Николаевич Колмогоров.

Далее он пишет: «Для успеха на олимпиаде необходимы некоторые специальные типы одарённости, которые вовсе не обязательны для успешной исследовательской работы. Уже само наличие назначенного очень ограниченного срока для решения задач многих делает совершенно беспомощными. Но существуют и такие математические проблемы, которые могут быть решены лишь в результате очень длительного и спокойного размышления и формирования новых понятий. Много такого рода проблем было решено замечательным советским топологом П. С. Александровым.» И далее — слушайте! «Не случайно Павел Сергеевич Александров говорил, что если бы во времена его юности были математические олимпиады, он, возможно, не сделался бы математиком: его главные достижения в математике явились не плодом быстро работающей *изобретательности,* а итогом длительного и углублённого *созерцания*». Прекрасно сказано!

Но вот те на! Вдруг выясняется, что гимн олимпиаде воспел чело-

век, которому повезло родиться при пресловутом царском режиме, когда таких олимпиад не было.

Но существует масса людей, которым повезло именно в том, что олимпиады были!

Посмотрим же, как олимпийский огонь освещал жизнь первым победителям московских олимпиад.

Победители и их судьбы

Кто же были они — победители и иные участники первых олимпиад? Как сложились их судьбы?

Снова предоставим слово Бончковскому.

«Победителями были признаны:

1) Зверев Игорь Николаевич (24 школа Дзержинского района).

2) Коробов Николай Михайлович (24 школа Бауманского района).

3) Мышкис Анна Вениаминовна (10 школа Краснопресненского района).

Все они были премированы небольшими математическими библиотечками».

... Все они поступили на мехмат. Зверев и Коробов потом всю жизнь работали и ныне работают на родном факультете. Аня Мышкис закончила факультет, потом — фронт, где она была связисткой. С фронта она не вернулась.

Из пяти человек, лауреатов второй премии, лишь «Джемс-Леви Юрий Евгеньевич (35 школа Краснопресненского района)» печатал в дальнейшем работы по математике. Он тоже был участником войны, был тяжело ранен, работал в Вычислительном центре Академии наук. И ещё (не названный в общем списке, но указанный на стр. 59 как автор решения одной из задач и лауреат второй премии) И. М. Кирко поступил на физфак, стал физиком, профессором, доктором физико-математических наук.

Вспомним: «Наконец, комитет рекомендовал участникам олимпиады посещать собрания Школьного математического кружка...»

Продолжим эту фразу отрывком из воспоминаний Якова Рафаиловича Бермана — профессора, доктора технических наук, двоюродного брата Бориса Владимировича Шабата (который был старостой этого кружка; Боря Шабат привлёк своего двоюродного брата к участию в кружке).

«Потом при Математическом институте им. Стеклова был создан математический кружок для московских школьников 9-10 классов (тогда этот институт находился $\langle\cdots\rangle$ на Большой Калужской улице, недалеко от теперешней улицы академика Петровского). $\langle\cdots\rangle$ Мы ездили в

Стекловку в так называемые общевыходные дни (6-е, 12-е, 18-е, 24-е и 30-е каждого месяца)[3]. Там мы слушали и Б. Н. Делоне, и Л. А. Люстерника, и Л. Г. Шнирельмана. А душой, руководителем кружка был двадцатидвухлетний преподаватель кафедры математического анализа мехмата МГУ И. М. Гельфанд. Среди кружковцев промелькнули Ю. Гермейер, А. Брудно, Н. Моисеев, О. Сорокин, Н. Коробов, ставшие впоследствии, как и Борис, студентами мехмата.»

... Вспоминает Анатолий Дмитриевич Мышкис (он тоже был участником кружка): «Гельфанд спросил, кто чем занимается. Многие ответили, что решают проблему Ферма. Гельфанд был полон иронии по этому поводу. А потом он стал рассказывать про многое, даже про логику, раздавал читать разные книжки...» А затем пошли лекции Люстерника, Колмогорова, не всегда понятные, но оставляющие впечатление чего-то великого.

И возжегся огонь. На всю жизнь.

Что же случилось с ними в дальнейшем?

Александр Львович Брудно — доктор физико-математических наук, профессор, работал в Институте электронных вычислительных машин, преподавал в педагогическом институте; Юрий Борисович Гермейер (1918 – 1975) — доктор физико-математических наук, профессор, работавший в Вычислительном центре АН СССР, заведовал кафедрой на факультете вычислительной математики и кибернетики МГУ; Никита Николаевич Моисеев — академик РАН, он был замдиректора Вычислительного Центра АН СССР, первым деканом факультета управления и прикладной математики МФТИ; Олег Сорокин, которого считали очень талантливым, погиб на фронте; о Коробове было уже сказано; ещё упоминался Борис Владимирович Шабат (1917-1987) — доктор физико-математических наук, он стал профессором мехмата МГУ.

Какой срез нашей истории просматривается сквозь судьбы этих людей!

Вот видится мне Александр Львович Брудно, одиноко сидящий на берегу Средиземного моря. Мне передали его шутку: «Здесь говорят на трёх языках: иврите, английском и русском. Как они не могут понять, что лучше всего говорить по-русски!»

Начало жизни Н. Н. Моисеева было трагическим. Как дворянин, он был изгоем. Одна из глав его воспоминаний так и названа «Изгой». Он сдал экзамены, но его не приняли на мехмат. Во время подготовки к экзаменам, друзья — Моисеев и Гермейер — помогали мальчику из бело-

[3] В конце двадцатых годов были отменены дни недели; их ввели обратно на памяти автора статьи, и дедушка с бабушкой, воспитывавшие меня, затруднились объяснить мне, что значит слово «воскресенье».

русской глубинки Семену Шапиро. Он получил много троек, в том числе по математике, но был зачислен. Это показалось юноше несправедливым, и он потащил Моисеева на прием к заместителю декана и начал громко и темпераментно возмущаться очевидной несправедливостью. Зам. декана его прервал: «Чего Вы хотите, Моисеев? Посмотрите на себя и на него (он пальцем показал на Семена) и подумайте, кого должно принять в университет рабоче-крестьянское государство?» Это был отказ.

Моисеев колебался, не стать ли ему тренером по лыжам (он был первоклассным лыжником). Всю зиму тренировался. Весной пришёл навестить своих товарищей по кружку. Встретил Гельфанда. Тот спросил: «Что-то я Вас не вижу, Моисеев. Как Вы сдали сессию?» «Я не был принят на мехмат», и он рассказал свою историю. Гельфанд отвёл юношу к декану — Л. А. Тумаркину, и сказал: «Примите его, он будет не хуже среднего студента». И Моисеев был принят.

... Через много лет Гельфанд и Моисеев одновременно стали академиками, и на банкете, посвящённом этому событию, вспоминали давний эпизод — поступление на мехмат.

Потом была Война. (Никита Николаевич пишет: «Свою общественную полноценность я впервые начал ощущать только во время войны».) Затем — долгая и плодотворная жизнь. Начиная с определённого момента, ему была оказана «полная и всесторонняя поддержка со стороны советского государства».

(А затем наступила перестройка, и «поддержка» прекратилась.)

Борис Владимирович Шабат пошёл в ополчение без ноги, на протезе! Чудом остался жив. Потом в его жизни было много и прекрасных, и нелегких моментов. И не всегда ощущалась «поддержка со стороны государства», бывало и наоборот. Но добрые люди обычно оказывались рядом, и трудности преодолевались. (А для многих и многих других, именно Борис Владимирович был этим «добрым человеком».)

Но у всех этих мальчиков, тех кто не погиб в Войну, было то, что невозможно было у них отнять: счастье учиться на лучшем в мире математическом факультете и любимая профессия. И этому во многом они обязаны кружку и олимпиаде.

Продолжим.

Вскоре был организован кружок в МГУ. Председателем бюро кружка 1935/36 года был Марк Глезерман. Тогда он был студентом 2 курса. Потом стал учеником Льва Семёновича Понтрягина. Написал совместно со своим научным руководителем посвящённую изложению двойственности Понтрягина статью в «Успехи математических наук» (опубликованную

после войны). Марк пошёл в ополчение. Попал в плен. Есть свидетельство, что при попытке к бегству он был схвачен и повешен.

Как пишут в своей замечательной статье «Школьный математический кружок при МГУ и московские математические олимпиады»[4)] Владимир Григорьевич Болтянский и Исаак Моисеевич Яглом (кстати сказать, победители, соответственно, шестой и четвёртой олимпиад), «решительная перестройка работы кружка связана с именем студента МГУ Додика Шклярского, талантливого математика и блестящего преподавателя, руководившего работой кружков в 1938-1941 годах.»

Шклярский получил первую премию на второй олимпиаде. Поступил на мехмат. В студенческие годы выполнил две работы, посвящённые анализу и топологии, опубликованные в 1944–45 годах. Закончил мехмат в 1941 году. Ушёл на фронт. Трагически погиб (подробности см. в «Общей газете», №18(197) от 8–14 мая 1997 года).

Друзья вспоминают, что Шклярский был фанатично предан математике. Он мог без конца говорить о ней. Очень любил возиться со школьниками. Он изменил стиль работы кружков. Заменил доклады школьников на решение задач. С тех пор так и повелось. Теперь такая форма кружковой работы и даже работы многих математических школ стала доминирующей.

«Достоинства новой системы,— пишут Болтянский и Яглом, — были проверены прямым экспериментом. В 1937/38 учебном году Шклярский проводил в своей секции занятия по описанной выше схеме, в то время как остальные секции работали по-старинке, главным образом ограничиваясь докладами школьников. Результат превзошёл все ожидания: на IV олимпиаде (1938 год) участники секции Шклярского унесли половину всех премий (12 из 24), в том числе все 4 первые премии! $\langle\cdots\rangle$ В числе последователей Шклярского отметим таких замечательных руководителей секций, как А. С. Кронрод, Е. Б. Дынкин, и воспитанных на традициях кружка В. И. Арнольда и А. А. Кириллова.»

С именем Евгения Борисовича Дынкина тоже связана целая эпоха в истории школьных кружков и московских математических олимпиад. Дынкинские школьные кружки плавно переходили в университетские семинары для первокурсников. (Среди участников такого семинара 1952 года был и автор этих строк.) Многие участники этих семинаров становились учениками Евгения Борисовича, затем — крупными исследователями.

[4)] Опубликованной в книге «Сборник задач московских олимпиад» / Сост. А. А. Леман. М.:Просвещение, 1965.

Сам Дынкин поступил на мехмат в 1940 году. Отец его был репрессирован и погиб в ГУЛАГе, и Евгений Борисович много раз говорил, что свое поступление на мехмат он воспринимал, как чудо (детей с такими анкетными данными обычно не брали). Он быстро проявил свою творческую незаурядность, и надо сказать, что «подрастающему таланту было обеспечено полное внимание, полная и всесторонняя помощь и поддержка», но не «со стороны советского государства и всего социалистического общества нашей страны», а со стороны очень доброго и отзывчивого человека — Софьи Александровны Яновской. А потом — Андрея Николаевича Колмогорова. В результате Е. Б. Дынкин стал профессором мехмата МГУ, замечательным математиком (в двух весьма разнородных областях математики — алгебре и теории вероятностей), основателем большой и плодотворно работающей научной школы. Однако в 1968 году он вынужден был уйти из Московского университета. В 1976 году Евгений Борисович эмигрирует в США. С 1977 года он — профессор Корнельского университета, член Американской академии наук и искусств и Национальной академии наук США.

... Несколько лет тому назад он приехал в Москву. Он пригласил нескольких своих бывших участников кружков к себе в гостиницу — пообщаться. Он прибыл в гостиницу с лекции для школьников, которую ему устроили его бывшие коллеги и ученики. В тот вечер он был окрылённым и преисполненным радости. Он увлёк школьников рассказом о старинной задаче («о разборчивой невесте», я слышал о ней от него в начале шестидесятых годов). Судя по всему, школьники слушали его с упоением, с горящими глазами, так же, как это было раньше, в далекие годы.

... В Америке Е. Б. Дынкин достиг очень многого: должности профессора одного из ведущих университетов, академических званий, получил кафедру, ранее занимаемую одним из крупнейших вероятностников нашего века (Ито). Он материально обеспечен, приобрел большой дом. Но рассказать школьникам о разборчивой невесте в Америке затруднительно: таких школьных кружков и олимпиад, как в России, в Америке нет. И какая-то часть души Евгения Борисовича оказалась там невостребованной.

Будем же гордиться тем, что родилось 63 года назад и пестовалось многими поколениями математиков — нашими школьными кружками и олимпиадами.

Я хотел бы выразить глубокую благодарность олимпийцам тридцатых годов: Я. Р. Берману, Л. И. Головиной, Н. М. Коробову, А. Д. Мышкису и П. Н. Папушу, поделившимися со мной своими воспоминаниями.

Тема номера: математика и криптография

Шифр, ключ, криптография... Эти слова из арсенала секретных служб и детективных романов уже давно перекочевали в нашу повседневную жизнь. Ведь информация в настоящее время хранится, обрабатывается и передаётся, в основном, с помощью электроники. Поэтому для обеспечения конфиденциальности приходится всё шире разрабатывать и применять средства защиты информации. Это вызвало рост интереса к криптографии как к науке о таких средствах. Раньше криптография была одной из самых секретных областей знаний и может быть поэтому в литературе на русском языке практически отсутствует общедоступное изложение научных основ криптографии. Зато в избытке имеются «околокриптографические» публикации: в них — легенды «вокруг криптографии», недобросовестная реклама криптографических продуктов и т.п. Вместе с тем в настоящее время ежегодно проводится несколько международных конференций по криптографии, издаётся несколько специализированных журналов, в некоторых университетах читаются курсы лекций.

Что же такое криптография и какое место в ней занимают математические методы? Для того, чтобы прояснить эти вопросы, редакция и решила сделать криптографию темой очередного номера. В предлагаемых статьях в строгой, но общедоступной форме вводятся все основные понятия криптографии и необходимые математические модели; некоторые из математических проблем криптографии не нашли отражения в этих статьях — разговор о таких проблемах впереди. Авторы статей — математики, которые занимаются различными вопросами криптографии в течение многих лет:

Ященко В. В., Варновский Н. П. — лаборатория МГУ по математическим проблемам криптографии;

Нестеренко Ю. В. — механико-математический факультет МГУ;

Кабатянский Г. А. — Институт проблем передачи информации РАН.

Редакция надеется, что эти статьи помогут читателю правильно ориентироваться в потоке публикаций по криптографии.

Основные понятия криптографии*

В. В. Ященко

Введение

Как передать нужную информацию нужному адресату в тайне от других? Каждый из читателей в разное время и с разными целями наверняка пытался решить для себя эту практическую задачу (для удобства дальнейших ссылок назовем ее «задача ТП», т. е. задача *ТайноПиси*). Выбрав подходящее решение, он, скорее всего, повторил изобретение одного из способов скрытой передачи информации, которым уже не одна тысяча лет.

Размышляя над задачей ТП, нетрудно прийти к выводу, что есть три возможности.

1. Создать абсолютно надежный, недоступный для других канал связи между абонентами.

2. Использовать общедоступный канал связи, но скрыть сам факт передачи информации.

3. Использовать общедоступный канал связи, но передавать по нему нужную информацию в так преобразованном виде, чтобы восстановить ее мог только адресат.

Прокомментируем эти три возможности.

1. При современном уровне развития науки и техники сделать такой канал связи между удаленными абонентами для неоднократной передачи больших объемов информации практически нереально.

2. Разработкой средств и методов скрытия факта передачи сообщения занимается *стеганография*.

Первые следы стеганографических методов теряются в глубокой древности. Например, известен такой способ скрытия письменного сообщения: голову раба брили, на коже головы писали сообщение и после отрастания волос раба отправляли к адресату.

Из детективных произведений хорошо известны различные способы тайнописи между строк обычного, незащищаемого текста: от молока до сложных химических реактивов с последующей обработкой.

*Настоящая статья является сокращенным переработанным вариантом книги С. А. Дориченко и В. В. Ященко «25 этюдов о шифрах», М.: Теис, 1994.

Также из детективов известен метод «*микроточки*»: сообщение записывается с помощью современной техники на очень маленький носитель (микроточку), который пересылается с обычным письмом, например, под маркой или где-нибудь в другом, заранее обусловленном месте.

В настоящее время в связи с широким распространением компьютеров известно много тонких методов «запрятывания» защищаемой информации внутри больших объемов информации, хранящейся в компьютере. Наглядный пример запрятывания текстового файла в графический можно найти в Интернете[1)]; он же приведен в журнале «Компьютерра», №48 (225) от 1 декабря 1997 г., на стр. 62. (Следует отметить, что авторы статьи в журнале ошибочно относят стеганографию к криптографии. Конечно, с помощью стеганографии можно прятать и предварительно зашифрованные тексты, но, вообще говоря, стеганография и криптография — принципиально различные направления в теории и практике защиты информации.)

3. Разработкой методов преобразования (*шифрования*) информации с целью ее защиты от незаконных пользователей занимается *криптография*. Такие методы и способы преобразования информации называются *шифрами*.

Шифрование (*зашифрование*) — процесс применения шифра к защищаемой информации, т. е. преобразование защищаемой информации (*открытого текста*) в шифрованное сообщение (*шифртекст, криптограмму*) с помощью определенных правил, содержащихся в шифре.

Дешифрование — процесс, обратный шифрованию, т. е. преобразование шифрованного сообщения в защищаемую информацию с помощью определенных правил, содержащихся в шифре.

Криптография — прикладная наука, она использует самые последние достижения фундаментальных наук и, в первую очередь, математики. С другой стороны, все конкретные задачи криптографии существенно зависят от уровня развития техники и технологии, от применяемых средств связи и способов передачи информации.

Предмет криптографии

Что же является предметом криптографии? Для ответа на этот вопрос вернемся к задаче ТП, чтобы уточнить ситуацию и используемые понятия.

Прежде всего заметим, что эта задача возникает только для информации, которая нуждается в защите. Обычно в таких случаях говорят, что информация содержит тайну или является *защищаемой, приватной, конфиденциальной, секретной*. Для наиболее типичных, часто встречаю-

[1)] `http://www.geocities.com/SiliconValley/Vista/6001/`

щихся ситуаций такого типа введены даже специальные понятия:
– государственная тайна;
– военная тайна;
– коммерческая тайна;
– юридическая тайна;
– врачебная тайна и т. д.

Далее мы будем говорить о защищаемой информации, имея в виду следующие признаки такой информации:

– имеется какой-то определенный круг *законных пользователей*, которые имеют право владеть этой информацией;

– имеются *незаконные пользователи*, которые стремятся овладеть этой информацией с тем, чтобы обратить ее себе во благо, а законным пользователям во вред.

Для простоты мы вначале ограничимся рассмотрением только одной *угрозы* — угрозы разглашения информации. Существуют и другие угрозы для защищаемой информации со стороны незаконных пользователей: подмена, имитация и др. О них мы поговорим ниже.

Теперь мы можем изобразить ситуацию, в которой возникает задача ТП, следующей схемой (см. рис. 1).

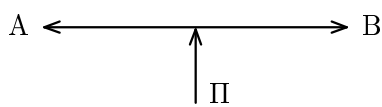


Рис. 1.

Здесь A и B — удаленные законные пользователи защищаемой информации; они хотят обмениваться информацией по общедоступному каналу связи. П — незаконный пользователь (*противник*), который может перехватывать передаваемые по каналу связи сообщения и пытаться извлечь из них интересующую его информацию. Эту формальную схему можно считать моделью типичной ситуации, в которой применяются криптографические методы защиты информации.

Отметим, что исторически в криптографии закрепились некоторые военные слова (противник, атака на шифр и др.) Они наиболее точно отражают смысл соответствующих криптографических понятий. Вместе с тем широко известная военная терминология, основанная на понятии кода (военно-морские коды, коды Генерального штаба, кодовые книги, кодобозначения и т. п.), уже не применяется в теоретической криптографии. Дело в том, что за последние десятилетия сформировалась *теория кодирования* — большое научное направление, которое разрабатывает и изучает методы защиты информации от случайных искажений в

каналах связи. И если ранее термины *кодирование* и *шифрование* употреблялись как синонимы, то теперь это недопустимо. Так, например, очень распространенное выражение «кодирование — разновидность шифрования» становится просто неправильным.

Криптография занимается методами преобразования информации, которые бы не позволили противнику извлечь ее из перехватываемых сообщений. При этом по каналу связи передается уже не сама защищаемая информация, а результат ее преобразования с помощью шифра, и для противника возникает сложная задача *вскрытия шифра*.

Вскрытие (*взламывание*) *шифра* — процесс получения защищаемой информации из шифрованного сообщения без знания примененного шифра.

Однако помимо перехвата и вскрытия шифра противник может пытаться получить защищаемую информацию многими другими способами. Наиболее известным из таких способов является агентурный, когда противник каким-либо путем склоняет к сотрудничеству одного из законных пользователей и с помощью этого агента получает доступ к защищаемой информации. В такой ситуации криптография бессильна.

Противник может пытаться не получить, а уничтожить или модифицировать защищаемую информацию в процессе ее передачи. Это — совсем другой тип угроз для информации, отличный от перехвата и вскрытия шифра. Для защиты от таких угроз разрабатываются свои специфические методы.

Следовательно, на пути от одного законного пользователя к другому информация должна защищаться различными способами, противостоящими различным угрозам. Возникает ситуация цепи из разнотипных звеньев, которая защищает информацию. Естественно, противник будет стремиться найти самое слабое звено, чтобы с наименьшими затратами добраться до информации. А значит, и законные пользователи должны учитывать это обстоятельство в своей стратегии защиты: бессмысленно делать какое-то звено очень прочным, если есть заведомо более слабые звенья («принцип равнопрочности защиты»).

Не следует забывать и ещё об одной важной проблеме: проблеме соотношения цены информации, затрат на ее защиту и затрат на ее добывание. При современном уровне развития техники сами средства связи, а также разработка средств перехвата информации из них и средств защиты информации требуют очень больших затрат. Прежде чем защищать информацию, задайте себе два вопроса:

1) является ли она для противника более ценной, чем стоимость атаки;

2) является ли она для вас более ценной, чем стоимость защиты.

Именно перечисленные соображения и являются решающими при выборе подходящих средств защиты: физических, стеганографических, криптографических и др.

Некоторые понятия криптографии удобно иллюстрировать историческими примерами, поэтому сделаем небольшое историческое отступление.

Долгое время занятие криптографией было уделом чудаков-одиночек. Среди них были одаренные учёные, дипломаты, священнослужители. Известны случаи, когда криптография считалась даже черной магией. Этот период развития криптографии как искусства длился с незапамятных времен до начала XX века, когда появились первые шифровальные машины. Понимание математического характера решаемых криптографией задач пришло только в середине XX века — после работ выдающегося американского учёного К. Шеннона.

История криптографии связана с большим количеством дипломатических и военных тайн и поэтому окутана туманом легенд. Наиболее полная книга по истории криптографии содержит более тысячи страниц. Она опубликована в 1967 году и на русский язык не переведена[2)]. На русском языке недавно вышел в свет фундаментальный труд по истории криптографии в России[3)].

Свой след в истории криптографии оставили многие хорошо известные исторические личности. Приведем несколько наиболее ярких примеров. Первые сведения об использовании шифров в военном деле связаны с именем спартанского полководца Лисандра (шифр «Сциталь»). Цезарь использовал в переписке шифр, который вошёл в историю как «шифр Цезаря». В древней Греции был изобретен вид шифра, который в дальнейшем стал называться «квадрат Полития». Одну из первых книг по криптографии написал аббат И. Трителий (1462–1516), живший в Германии. В 1566 году известный математик Д. Кардано опубликовал работу с описанием изобретенной им системы шифрования («решётка Кардано»). Франция XVI века оставила в истории криптографии шифры короля Генриха IV и Ришелье. В упомянутой книге Т. А. Соболевой подробно описано много российских шифров, в том числе и «цифирная азбука» 1700 года, автором которой был Петр Великий.

Некоторые сведения о свойствах шифров и их применении можно найти и в художественной литературе, особенно в приключенческой, детективной и военной. Хорошее подробное объяснение особенностей одного из простейших шифров — *шифра замены* и методов его вскрытия содержится в двух известных рассказах: «Золотой жук» Э. По и «Пляшущие человечки» А. Конан-Дойля.

Рассмотрим более подробно два примера.

Шифр «Сциталь». Этот шифр известен со времен войны Спарты против Афин в V веке до н.э. Для его реализации использовался сциталь — жезл, имеющий форму цилиндра. На сциталь виток к витку наматывалась узкая папирусная лента (без пробелов и нахлестов), а затем на этой ленте вдоль оси сциталя записывался открытый текст. Лента разматывалась и получалось (для

[2)] *Kahn David.* Codebreakers. The story of Secret Writing. New York: Macmillan, 1967.

[3)] *Соболева Т. А.* Тайнопись в истории России (История криптографической службы России XVIII – начала XX в.). М., 1994.

непосвященных), что поперек ленты в беспорядке написаны какие-то буквы. Затем лента отправлялась адресату. Адресат брал такой же сциталь, таким же образом наматывал на него полученную ленту и читал сообщение вдоль оси сциталя.

Отметим, что в этом шифре преобразование открытого текста в шифрованный заключается в определенной перестановке букв открытого текста. Поэтому класс шифров, к которым относится и шифр «Сциталь», называется *шифрами перестановки*.

Шифр Цезаря. Этот шифр реализует следующее преобразование открытого текста: каждая буква открытого текста заменяется третьей после нее буквой в алфавите, который считается написанным по кругу, т. е. после буквы «я» следует буква «а». Отметим, что Цезарь заменял букву третьей после нее буквой, но можно заменять и какой-нибудь другой. Главное, чтобы тот, кому посылается шифрованное сообщение, знал эту величину сдвига. Класс шифров, к которым относится и шифр Цезаря, называется *шифрами замены*.

Из предыдущего изложения понятно, что придумывание хорошего шифра — дело трудоемкое. Поэтому желательно увеличить «время жизни» хорошего шифра и использовать его для шифрования как можно большего количества сообщений. Но при этом возникает опасность, что противник уже разгадал (вскрыл) шифр и читает защищаемую информацию. Если же в шифре есть сменный ключ, то, заменив ключ, можно сделать так, что разработанные противником методы уже не дают эффекта.

Под *ключом* в криптографии понимают сменный элемент шифра, который применяется для шифрования конкретного сообщения. Например, в шифре «Сциталь» ключом является диаметр сциталя, а в шифрах типа шифра Цезаря ключом является величина сдвига букв шифртекста относительно букв открытого текста.

Описанные соображения привели к тому, что безопасность защищаемой информации стала определяться в первую очередь ключом. Сам шифр, шифрмашина или принцип шифрования стали считать известными противнику и доступными для предварительного изучения, но в них появился неизвестный для противника ключ, от которого существенно зависят применяемые преобразования информации. Теперь законные пользователи, прежде чем обмениваться шифрованными сообщениями, должны тайно от противника обменяться ключами или установить одинаковый ключ на обоих концах канала связи. А для противника появилась новая задача — определить ключ, после чего можно легко прочитать зашифрованные на этом ключе сообщения.

Вернемся к формальному описанию основного объекта криптографии (рис. 1, стр. 55). Теперь в него необходимо внести существенное изменение — добавить недоступный для противника секретный канал связи для

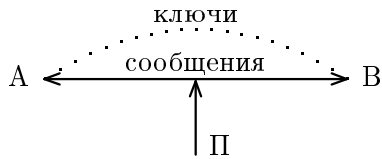


Рис. 2.

обмена ключами (см. рис. 2). Создать такой канал связи вполне реально, поскольку нагрузка на него, вообще говоря, небольшая.

Отметим теперь, что не существует единого шифра, подходящего для всех случаев. Выбор способа шифрования зависит от особенностей информации, ее ценности и возможностей владельцев по защите своей информации. Прежде всего подчеркнём большое разнообразие видов защищаемой информации: документальная, телефонная, телевизионная, компьютерная и т. д. Каждый вид информации имеет свои специфические особенности, и эти особенности сильно влияют на выбор методов шифрования информации. Большое значение имеют объемы и требуемая скорость передачи шифрованной информации. Выбор вида шифра и его параметров существенно зависит от характера защищаемых секретов или тайны. Некоторые тайны (например, государственные, военные и др.) должны сохраняться десятилетиями, а некоторые (например, биржевые) — уже через несколько часов можно разгласить. Необходимо учитывать также и возможности того противника, от которого защищается данная информация. Одно дело — противостоять одиночке или даже банде уголовников, а другое дело — мощной государственной структуре.

Способность шифра противостоять всевозможным атакам на него называют *стойкостью шифра*.

Под *атакой на шифр* понимают попытку вскрытия этого шифра.

Понятие стойкости шифра является центральным для криптографии. Хотя качественно понять его довольно легко, но получение строгих доказуемых оценок стойкости для каждого конкретного шифра — проблема нерешенная. Это объясняется тем, что до сих пор нет необходимых для решения такой проблемы математических результатов. (Мы вернемся к обсуждению этого вопроса ниже.) Поэтому стойкость конкретного шифра оценивается только путем всевозможных попыток его вскрытия и зависит от квалификации *криптоаналитиков*, атакующих шифр. Такую процедуру иногда называют *проверкой стойкости*.

Важным подготовительным этапом для проверки стойкости шифра является продумывание различных предполагаемых возможностей, с

помощью которых противник может атаковать шифр. Появление таких возможностей у противника обычно не зависит от криптографии, это является некоторой внешней подсказкой и существенно влияет на стойкость шифра. Поэтому оценки стойкости шифра всегда содержат те предположения о целях и возможностях противника, в условиях которых эти оценки получены.

Прежде всего, как это уже отмечалось выше, обычно считается, что противник знает сам шифр и имеет возможности для его предварительного изучения. Противник также знает некоторые характеристики открытых текстов, например, общую тематику сообщений, их стиль, некоторые стандарты, форматы и т. д.

Из более специфических приведем ещё три примера возможностей противника:
– противник может перехватывать все шифрованные сообщения, но не имеет соответствующих им открытых текстов;
– противник может перехватывать все шифрованные сообщения и добывать соответствующие им открытые тексты;
– противник имеет доступ к шифру (но не к ключам!) и поэтому может зашифровывать и дешифровывать любую информацию.

На протяжении многих веков среди специалистов не утихали споры о стойкости шифров и о возможности построения абсолютно стойкого шифра. Приведем три характерных высказывания на этот счёт.

Английский математик Чарльз Беббидж (XIX в.): «Всякий человек, даже если он не знаком с техникой вскрытия шифров, твердо считает, что сможет изобрести абсолютно стойкий шифр, и чем более умен и образован этот человек, тем более твердо это убеждение. Я сам разделял эту уверенность в течение многих лет.»

«Отец кибернетики» Норберт Винер: «Любой шифр может быть вскрыт, если только в этом есть настоятельная необходимость и информация, которую предполагается получить, стоит затраченных средств, усилий и времени...»

Автор шифра PGP Ф. Зиммерманн («Компьютерра», №48 от 1.12.1997, стр. 45–46):

«Каждый, кто думает, что изобрел непробиваемую схему шифрования, — или невероятно редкий гений, или просто наивен и неопытен...»

«Каждый программист воображает себя криптографом, что ведёт к распространению исключительно плохого криптообеспечения...»

В заключение данного раздела сделаем ещё одно замечание — о терминологии. В последнее время наряду со словом «криптография» часто встречается и слово «криптология», но соотношение между ними не всегда понимается правильно. Сейчас происходит окончательное формирование этих научных дисциплин, уточняются их предмет и задачи.

Криптология — наука, состоящая из двух ветвей: криптографии и криптоанализа.

Криптография — наука о способах преобразования (шифрования) информации с целью ее защиты от незаконных пользователей.

Криптоанализ — наука (и практика ее применения) о методах и способах вскрытия шифров.

Соотношение криптографии и криптоанализа очевидно: криптография — защита, т. е. разработка шифров, а криптоанализ — нападение, т. е. атака на шифры. Однако эти две дисциплины связаны друг с другом, и не бывает хороших криптографов, не владеющих методами криптоанализа.

Математические основы

Большое влияние на развитие криптографии оказали появившиеся в середине нашего века работы американского математика Клода Шеннона. В этих работах были заложены основы теории информации, а также был разработан математический аппарат для исследований во многих областях науки, связанных с информацией. (Отметим, кстати, что нынешний 1998 год — юбилейный для теории информации. Принято считать, что теория информации как наука родилась в 1948 году после публикации работы К. Шеннона «Математическая теория связи»[4].)

В своей работе «Теория связи в секретных системах» Клод Шеннон обобщил накопленный до него опыт разработки шифров. Оказалось, что даже в сложных шифрах в качестве типичных компонентов можно выделить *шифры замены, шифры перестановки* или их сочетания.

Шифр замены является простейшим, наиболее популярным шифром. Типичными примерами являются шифр Цезаря, «цифирная азбука» Петра Великого и «пляшущие человечки» А. Конан-Дойля. Как видно из самого названия, шифр замены осуществляет преобразование замены букв или других «частей» открытого текста на аналогичные «части» шифрованного текста. Легко дать математическое описание шифра замены. Пусть X и Y — два алфавита (открытого и шифрованного текстов соответственно), состоящие из одинакового числа символов. Пусть также $g\colon X \to Y$ — взаимнооднозначное отображение X в Y. Тогда шифр замены действует так: открытый текст $x_1x_2\ldots x_n$ преобразуется в шифрованный текст $g(x_1)g(x_2)\ldots g(x_n)$.

[4] *Shannon C. E.* A mathematical theory of communication // Bell System Techn. J. V. 27, №3, 1948. P. 379–423; V. 27, №4, 1948. P. 623–656.

Шифр перестановки, как видно из названия, осуществляет преобразование перестановки букв в открытом тексте. Типичным примером шифра перестановки является шифр «Сциталь». Обычно открытый текст разбивается на отрезки равной длины и каждый отрезок шифруется независимо. Пусть, например, длина отрезков равна n и σ — взаимнооднозначное отображение множества $\{1, 2, \ldots, n\}$ в себя. Тогда шифр перестановки действует так: отрезок открытого текста $x_1 \ldots x_n$ преобразуется в отрезок шифрованного текста $x_{\sigma(1)} \ldots x_{\sigma(n)}$.

Важнейшим для развития криптографии был результат К. Шеннона о существовании и единственности абсолютно стойкого шифра. Единственным таким шифром является какая-нибудь форма так называемой *ленты однократного использования*, в которой открытый текст «объединяется» с полностью случайным ключом такой же длины.

Этот результат был доказан К. Шенноном с помощью разработанного им теоретико-информационного метода исследования шифров. Мы не будем здесь останавливаться на этом подробно, заинтересованному читателю рекомендуем изучить работу К. Шеннона[5)].

Обсудим особенности строения абсолютно стойкого шифра и возможности его практического использования. Типичным и наиболее простым примером реализации абсолютно стойкого шифра является шифр Вернама, который осуществляет побитовое сложение n-битового открытого текста и n-битового ключа:

$$y_i = x_i \oplus k_i, \quad i = 1, \ldots, n.$$

Здесь $x_1 \ldots x_n$ — открытый текст, $k_1, \ldots, k_n$ — ключ, $y_1 \ldots y_n$ — шифрованный текст.

Подчеркнём, что для абсолютной стойкости существенным является каждое из следующих требований к ленте однократного использования:

1) полная случайность (равновероятность) ключа (это, в частности, означает, что ключ нельзя вырабатывать с помощью какого-либо детерминированного устройства);

2) равенство длины ключа и длины открытого текста;

3) однократность использования ключа.

В случае нарушения хотя бы одного из этих условий шифр перестает быть абсолютно стойким и появляются принципиальные возможности для его вскрытия (хотя они могут быть трудно реализуемыми).

[5)] *Shannon C. E.* Communication theory of secrecy systems // Bell System Techn. J. V. 28, №4, 1949. P. 656–715.
Русск. пер. в: *Шеннон К.* Работы по теории информации и кибернетике. М.: ИЛ, 1963. С. 333–403.

Но, оказывается, именно эти условия и делают абсолютно стойкий шифр очень дорогим и непрактичным. Прежде чем пользоваться таким шифром, мы должны обеспечить всех абонентов достаточным запасом случайных ключей и исключить возможность их повторного применения. А это сделать необычайно трудно и дорого.

Как отмечал Д. Кан: «Проблема создания, регистрации, распространения и отмены ключей может показаться не слишком сложной тому, кто не имеет опыта передачи сообщений по каналам военной связи, но в военное время объем передаваемых сообщений ставит в тупик даже профессиональных связистов. За сутки могут быть зашифрованы сотни тысяч слов. Создание миллионов ключевых знаков потребовало бы огромных финансовых издержек и было бы сопряжено с большими затратами времени. Так как каждый текст должен иметь свой собственный, единственный и неповторимый ключ, применение идеальной системы потребовало бы передачи по крайней мере такого количества знаков, которое эквивалентно всему объему передаваемой военной информации.»

В силу указанных причин абсолютно стойкие шифры применяются только в сетях связи с небольшим объемом передаваемой информации, обычно это сети для передачи особо важной государственной информации.

Теперь уже понятно, что чаще всего для защиты своей информации законные пользователи вынуждены применять неабсолютно стойкие шифры. Такие шифры, по крайней мере теоретически, могут быть вскрыты. Вопрос только в том, хватит ли у противника сил, средств и времени для разработки и реализации соответствующих алгоритмов. Обычно эту мысль выражают так: противник с неограниченными ресурсами может вскрыть любой неабсолютно стойкий шифр.

Как же должен действовать в этой ситуации законный пользователь, выбирая для себя шифр? Лучше всего, конечно, было бы доказать, что никакой противник не может вскрыть выбранный шифр, скажем, за 10 лет и тем самым получить теоретическую оценку стойкости. К сожалению, математическая теория ещё не даёт нужных теорем — они относятся к нерешенной *проблеме нижних оценок сложности задач*.

Поэтому у пользователя остается единственный путь — получение практических оценок стойкости. Этот путь состоит из следующих этапов:

– понять и чётко сформулировать, от какого противника мы собираемся защищать информацию; необходимо уяснить, что именно противник знает или сможет узнать о системе шифра, а также какие силы и средства он сможет применить для его вскрытия;

– мысленно стать в положение противника и пытаться с его позиций атаковать шифр, т. е. разрабатывать различные алгоритмы вскрытия

шифра; при этом необходимо в максимальной мере обеспечить моделирование сил, средств и возможностей противника;
– наилучший из разработанных алгоритмов использовать для практической оценки стойкости шифра.

Здесь полезно для иллюстрации упомянуть о двух простейших методах вскрытия шифра: случайное угадывание ключа (он срабатывает с маленькой вероятностью, зато имеет маленькую сложность) и перебор всех подряд ключей вплоть до нахождения истинного (он срабатывает всегда, зато имеет очень большую сложность). Отметим также, что не всегда нужна атака на ключ: для некоторых шифров можно сразу, даже не зная ключа, восстанавливать открытый текст по шифрованному.

Новые направления

В 1983 году в книге «Коды и математика» М. Н. Аршинова и Л. Е. Садовского (библиотечка «Квант») было написано: «Приемов тайнописи — великое множество, и, скорее всего, это та область, где уже нет нужды придумывать что-нибудь существенно новое.» Однако это было очередное большое заблуждение относительно криптографии. Еще в 1976 году была опубликована работа молодых американских математиков У. Диффи и М. Э. Хеллмана «Новые направления в криптографии»[6)], которая не только существенно изменила криптографию, но и привела к появлению и бурному развитию новых направлений в математике. Центральным понятием «новой криптографии» является понятие односторонней функции (подробнее об этом см. статью Н. П. Варновского в настоящем номере журнала, стр. 71–86).

Односторонней называется функция $F\colon X \to Y$, обладающая двумя свойствами:
а) существует полиномиальный алгоритм вычисления значений $F(x)$;
б) не существует полиномиального алгоритма *инвертирования* функции F (т. е. решения уравнения $F(x) = y$ относительно x для непренебрежимо малой доли области значений функции).

Отметим, что односторонняя функция существенно отличается от функций, привычных со школьной скамьи, из-за ограничений на сложность ее вычисления и инвертирования. Вопрос о существовании односторонних функций пока открыт.

Другим понятием, более близким к традиционной криптографии, в которой есть секретный ключ, является понятие *функции с секретом*.

[6)] *Диффи У., Хеллман М. Э.* Защищенность и имитостойкость. Введение в криптографию // ТИИЭР. Т. 67, №3, 1979.

Иногда ещё употребляется термин *функция с ловушкой*. *Функцией с секретом* K называется функция $F_K\colon X \to Y$, зависящая от параметра K и обладающая тремя свойствами:
а) при любом K существует полиномиальный алгоритм вычисления значений $F_K(x)$;
б) при неизвестном K не существует полиномиального алгоритма инвертирования F_K;
в) при известном K существует полиномиальный алгоритм инвертирования F_K.

Про существование функций с секретом можно сказать то же самое, что сказано про односторонние функции. Для практических целей криптографии было построено несколько функций, которые могут оказаться функциями с секретом. Для них свойство б) пока строго не доказано, но известно, что задача инвертирования эквивалентна некоторой давно изучаемой трудной математической задаче. Наиболее известной и популярной из них является теоретико-числовая функция, на которой построен шифр RSA (подробнее об этом см. статью Ю. В. Нестеренко в настоящем номере журнала, стр. 87–114).

Применение функций с секретом в криптографии позволяет:

1) организовать обмен шифрованными сообщениями с использованием только открытых каналов связи, т. е. отказаться от секретных каналов связи для предварительного обмена ключами;

2) включить в задачу вскрытия шифра трудную математическую задачу и тем самым повысить обоснованность стойкости шифра;

3) решать новые криптографические задачи, отличные от шифрования (*цифровая подпись* и др.).

Опишем, например, как можно реализовать п. 1). Пользователь A, который хочет получать шифрованные сообщения, должен выбрать какую-нибудь функцию F_K с секретом K. Он сообщает всем заинтересованным (например, публикует) описание функции F_K в качестве своего алгоритма шифрования. Но при этом значение секрета K он никому не сообщает и держит в секрете. Если теперь пользователь B хочет послать пользователю A защищаемую информацию $x \in X$, то он вычисляет $y = F_K(x)$ и посылает y по открытому каналу пользователю A. Поскольку A для своего секрета K умеет инвертировать F_K, то он вычисляет x по полученному y. Никто другой не знает K и поэтому в силу свойства б) функции с секретом не сможет за полиномиальное время по известному шифрованному сообщению $F_K(x)$ вычислить защищаемую информацию x.

Описанную систему называют *криптосистемой с открытым ключом*, поскольку алгоритм шифрования F_K является общедоступным или

открытым. В последнее время такие криптосистемы ещё называют *асимметричными*, поскольку в них есть асимметрия в алгоритмах: алгоритмы шифрования и дешифрования различны. В отличие от таких систем традиционные шифры называют *симметричными*: в них ключ для шифрования и дешифрования один и тот же, и именно поэтому его нужно хранить в секрете. Для асимметричных систем алгоритм шифрования общеизвестен, но восстановить по нему алгоритм дешифрования за полиномиальное время невозможно.

Описанную выше идею Диффи и Хеллман предложили использовать также для цифровой подписи сообщений, которую невозможно подделать за полиномиальное время. Пусть пользователю A необходимо подписать сообщение x. Он, зная секрет K, находит такое y, что $F_K(y) = x$, и посылает y пользователю B в качестве своей цифровой подписи. Пользователь B хранит y в качестве доказательства того, что A подписал сообщение x.

Сообщение, подписанное цифровой подписью, можно представлять себе как пару (x, y), где x — сообщение, y — решение уравнения $F_K(y) = x$, $F_K\colon X \to Y$ — функция с секретом, известная всем взаимодействующим абонентам. Из определения функции F_K очевидны следующие достоинства цифровой подписи:

1) подписать сообщение x, т. е. решить уравнение $F_K(y) = x$, может только абонент — обладатель данного секрета K; другими словами, подделать подпись невозможно;

2) проверить подлинность подписи может любой абонент, знающий открытый ключ, т. е. саму функцию F_K;

3) при возникновении споров отказаться от подписи невозможно в силу ее неподделываемости;

4) подписанные сообщения (x, y) можно, не опасаясь ущерба, пересылать по любым каналам связи.

Кроме принципа построения криптосистемы с открытым ключом, Диффи и Хеллман в той же работе предложили ещё одну новую идею — *открытое распределение ключей*. Они задались вопросом: можно ли организовать такую процедуру взаимодействия абонентов A и B по открытым каналам связи, чтобы решить следующие задачи:

1) вначале у A и B нет никакой общей секретной информации, но в конце процедуры такая общая секретная информация (общий ключ) у A и B появляется, т. е. вырабатывается;

2) противник, который перехватывает все передачи информации и знает, что хотят получить A и B, тем не менее не может восстановить выработанный общий ключ A и B.

Диффи и Хеллман предложили решать эти задачи с помощью функции

$$F(x) = \alpha^x \pmod{p},$$

где p — большое простое число, x — произвольное натуральное число, α — некоторый *примитивный элемент* поля $GF(p)$. Общепризнанно, что инвертирование функции $\alpha^x \pmod{p}$, т. е. дискретное логарифмирование, является трудной математической задачей.

Сама процедура или, как принято говорить, *протокол выработки общего ключа* описывается следующим образом.

Числа p и α считаются общедоступными.

Абоненты A и B независимо друг от друга случайно выбирают по одному натуральному числу — скажем x_A и x_B. Эти элементы они держат в секрете. Далее каждый из них вычисляет новый элемент:

$$y_A = \alpha^{x_A} \pmod{p}, \quad y_B = \alpha^{x_B} \pmod{p}.$$

Потом они обмениваются этими элементами по каналу связи. Теперь абонент A, получив y_B и зная свой секретный элемент x_A, вычисляет новый элемент

$$y_B^{x_A} \pmod{p} = (\alpha^{x_B})^{x_A} \pmod{p}.$$

Аналогично поступает абонент B:

$$y_A^{x_B} \pmod{p} = (\alpha^{x_A})^{x_B} \pmod{p}.$$

Тем самым у A и B появился общий элемент поля, равный $\alpha^{x_A x_B}$. Этот элемент и объявляется общим ключом A и B.

Из описания протокола видно, что противник знает $p, \alpha, \alpha^{x_A}, \alpha^{x_B}$, не знает x_A и x_B и хочет узнать $\alpha^{x_A x_B}$. В настоящее время нет алгоритмов действий противника, более эффективных, чем дискретное логарифмирование, а это — трудная математическая задача.

Успехи, достигнутые в разработке схем цифровой подписи и открытого распределения ключей, позволили применить эти идеи также и к другим задачам взаимодействия удаленных абонентов. Так возникло большое новое направление теоретической криптографии — криптографические протоколы. Под *криптографическим протоколом* понимают такую процедуру взаимодействия абонентов, в результате которой абоненты (не противники!) достигают своей цели, а противник — не достигает.

Объектом изучения теории криптографических протоколов являются удаленные абоненты, взаимодействующие по открытым каналам связи. Целью взаимодействия абонентов является решение какой-то задачи. Имеется также противник, который преследует собственные цели. При этом противник в разных задачах может иметь разные возможности: например, может взаимодействовать с абонентами от имени других абонентов или вмешиваться в обмены информацией между абонентами и т. д.

Противником может даже оказаться один из абонентов или несколько абонентов, вступивших в сговор.

Приведем ещё несколько примеров задач, решаемых удаленными абонентами. (Читателю рекомендуем по своему вкусу самостоятельно придумать ещё какие-нибудь примеры.)

1. Взаимодействуют два не доверяющих друг другу абонента. Они хотят подписать контракт. Это надо сделать так, чтобы не допустить следующую ситуацию: один из абонентов получил подпись другого, а сам не подписался.

Протокол решения этой задачи принято называть *протоколом подписания контракта.*

2. Взаимодействуют два не доверяющих друг другу абонента. Они хотят бросить жребий с помощью монеты. Это надо сделать так, чтобы абонент, подбрасывающий монету, не мог изменить результат подбрасывания после получения догадки от абонента, угадывающего этот результат.

Протокол решения этой задачи принято называть *протоколом подбрасывания монеты.*

Опишем один из простейших протоколов подбрасывания монеты по телефону (так называемая схема Блюма-Микали). Для его реализации у абонентов A и B должна быть односторонняя функция $f\colon X \to Y$, удовлетворяющая следующим условиям:

1) X — конечное множество целых чисел, которое содержит одинаковое количество чётных и нечётных чисел;

2) любые числа $x_1, x_2 \in X$, имеющие один образ $f(x_1) = f(x_2)$, имеют одну чётность;

3) по заданному образу $f(x)$ «трудно» вычислить чётность неизвестного аргумента x.

Роль подбрасывания монеты играет случайный и равновероятный выбор элемента $x \in X$, а роль орла и решки — чётность и нечётность x соответственно. Пусть A — абонент, подбрасывающий монету, а B — абонент, угадывающий результат. Протокол состоит из следующих шагов:

1) A выбирает x («подбрасывает монету»), зашифровывает x, т. е. вычисляет $y = f(x)$, и посылает y абоненту B;

2) B получает y, пытается угадать чётность x и посылает свою догадку абоненту A;

3) A получает догадку от B и сообщает B, угадал ли он, посылая ему выбранное число x;

4) B проверяет, не обманывает ли A, вычисляя значение $f(x)$ и сравнивая его с полученным на втором шаге значением y.

Рекомендуем читателю самостоятельно проверить, что необходимые требования к протоколу подбрасывания монеты выполнены из-за свойств функции f.

3. Взаимодействуют два абонента A и B (типичный пример: A — клиент банка, B — банк). Абонент A хочет доказать абоненту B, что он именно A, а не противник.

Протокол решения этой задачи принято называть *протоколом идентификации абонента.*

4. Взаимодействуют несколько удаленных абонентов, получивших приказы из одного центра. Часть абонентов, включая центр, могут быть противниками. Необходимо выработать единую стратегию действий, выигрышную для абонентов.

Эту задачу принято называть задачей о византийских генералах, а протокол ее решения — *протоколом византийского соглашения.*

Опишем пример, которому эта задача обязана своим названием. Византия. Ночь перед великой битвой. Византийская армия состоит из n легионов, каждый из которых подчиняется своему генералу. Кроме того, у армии есть главнокомандующий, который руководит генералами. Однако империя находится в упадке и до одной трети генералов, включая главнокомандующего, могут быть предателями. В течение ночи каждый из генералов получает от главнокомандующего приказ о действиях на утро, причем возможны два варианта приказа: «атаковать» или «отступать». Если все честные генералы атакуют, то они побеждают. Если все они отступают, то им удаётся сохранить армию. Но если часть из них атакует, а часть отступает, то они терпят поражение. Если главнокомандующий окажется предателем, то он может дать разным генералам разные приказы, поэтому приказы главнокомандующего не стоит выполнять беспрекословно. Если каждый генерал будет действовать независимо от остальных, результаты могут оказаться плачевными. Очевидно, что генералы нуждаются в обмене информацией друг с другом (относительно полученных приказов) с тем, чтобы прийти к соглашению.

Осмысление различных протоколов и методов их построения привело в 1985–1986 г.г. к появлению двух плодотворных математических моделей — *интерактивной системы доказательства* и *доказательства с нулевым разглашением.* Математические исследования этих новых объектов позволили доказать несколько утверждений, весьма полезных при разработке криптографических протоколов (подробнее об этом см. статью Н. П. Варновского в настоящем номере журнала, стр. 71–86).

Под интерактивной системой доказательства (P, V, S) понимают протокол взаимодействия двух абонентов: P (доказывающий) и V (проверяющий). Абонент P хочет доказать V, что утверждение S истинно. При этом абонент V самостоятельно, без помощи P, не может доказать утверждение S (поэтому V и называется проверяющим). Абонент P может быть и противником, который хочет доказать V, что утверждение S истинно, хотя оно ложно. Протокол может состоять из многих *раундов*

обмена сообщениями между P и V и должен удовлетворять двум условиям:
1) *полнота* — если S действительно истинно, то абонент P убедит абонента V признать это;
2) *корректность* — если S ложно, то абонент P вряд ли убедит абонента V, что S истинно.

Здесь словами «вряд ли» мы для простоты заменили точную математическую формулировку.

Подчеркнём, что в определении системы (P, V, S) не допускалось, что V может быть противником. А если V оказался противником, который хочет «выведать» у P какую-нибудь новую полезную для себя информацию об утверждении S? В этом случае P, естественно, может не хотеть, чтобы это случилось в результате работы протокола (P, V, S). Протокол (P, V, S), решающий такую задачу, называется доказательством с нулевым разглашением и должен удовлетворять, кроме условий 1) и 2), ещё и следующему условию:

3) *нулевое разглашение* — в результате работы протокола (P, V, S) абонент V не увеличит свои знания об утверждении S или, другими словами, не сможет извлечь никакой информации о том, почему S истинно.

Заключение

За последние годы криптография и криптографические методы всё шире входят в нашу жизнь и даже быт. Вот несколько примеров. Отправляя Email, мы в некоторых случаях отвечаем на вопрос меню: «Нужен ли режим зашифрования?» Владелец интеллектуальной банковской карточки, обращаясь через терминал к банку, вначале выполняет криптографический протокол аутентификации карточки. Пользователи сети Интернет наверняка знакомы с дискуссиями вокруг возможного принятия стандарта цифровой подписи для тех страниц, которые содержат «критическую » информацию (юридическую, прайс-листы и др.). С недавних пор пользователи сетей стали указывать после своей фамилии наряду с уже привычным «Email . . . » и менее привычное — «Отпечаток открытого ключа . . . ».

С каждым днем таких примеров становится всё больше. Именно новые практические приложения криптографии и являются одним из источников ее развития. Криптографии, как и любой другой науке, необходимы новые нетривиальные и неожиданные идеи. Автор настоящей статьи надеется, что кто-то из ее читателей станет автором новых идей, а, быть может, и новейших направлений в криптографии.

Криптография и теория сложности

Н. П. Варновский

В небольшой по объему журнальной статье невозможно дать систематическое изложение основ какой-либо математической теории. Поэтому основное внимание мы уделяем разъяснению важнейших идей, связанных с применением теоретико-сложностного подхода в криптографии. Изложение по необходимости недостаточно формальное — для математической криптографии типичны многостраничные определения. Предполагается знакомство читателя с основами теории сложности вычислений: понятиями машины Тьюринга, классов P и NP (см. [2]), а также со статьей В. В. Ященко в настоящем номере журнала, стр. 53–70.

1. Введение

В теоретической криптографии существуют два основных подхода к определению стойкости криптосистем и криптографических протоколов (в дальнейшем мы будем также использовать общий термин — криптографические схемы): теоретико-информационный и теоретико-сложностной. Теоретико-информационный подход предполагает, что противник, атакующий криптографическую схему, не имеет даже теоретической возможности получить информацию, достаточную для осуществления своих целей. Классическим примером здесь может служить шифр Вернама с одноразовыми ключами, абсолютно стойкий против пассивного противника.

Подавляющее большинство используемых на практике криптографических схем не обладает столь высокой стойкостью. Более того, обычно бывает несложно указать алгоритм, который выполняет стоящую перед противником задачу, но не практически, а в принципе. Рассмотрим следующий пример.

Пример 1 (Криптосистема с открытым ключом). *Криптосистема с открытым ключом* полностью определяется тремя алгоритмами: генерации ключей, шифрования и дешифрования. Алгоритм генерации ключей G общедоступен; всякий желающий может подать ему на вход случайную строку r надлежащей длины и получить пару ключей (K_1, K_2).

Открытый ключ K_1 публикуется, а секретный ключ K_2 и случайная строка r хранятся в секрете. Алгоритмы шифрования E_{K_1} и дешифрования D_{K_2} таковы, что если (K_1, K_2) — пара ключей, сгенерированная алгоритмом G, то $D_{K_2}(E_{K_1}(m)) = m$ для любого открытого текста m. Для простоты изложения предполагаем, что открытый текст и криптограмма имеют одинаковую длину n. Кроме того, считаем, что открытый текст, криптограмма и оба ключа являются строками в двоичном алфавите.

Предположим теперь, что противник атакует эту криптосистему. Ему известен открытый ключ K_1, но неизвестен соответствующий секретный ключ K_2. Противник перехватил криптограмму d и пытается найти сообщение m, где $d = E_{K_1}(m)$. Поскольку алгоритм шифрования общеизвестен, противник может просто последовательно перебрать все возможные сообщения длины n, вычислить для каждого такого сообщения m_i криптограмму $d_i = E_{K_1}(m_i)$ и сравнить d_i с d. То сообщение, для которого $d_i = d$, и будет искомым открытым текстом. Если повезет, то открытый текст будет найден достаточно быстро. В худшем же случае перебор будет выполнен за время порядка $2^n T(n)$, где $T(n)$ — время, требуемое для вычисления функции E_{K_1} от сообщений длины n. Если сообщения имеют длину порядка 1000 битов, то такой перебор неосуществим на практике ни на каких самых мощных компьютерах.

Мы рассмотрели лишь один из возможных способов атаки на криптосистему и простейший алгоритм поиска открытого текста, называемый обычно алгоритмом полного перебора. Используется также и другое название: «метод грубой силы». Другой простейший алгоритм поиска открытого текста — угадывание. Этот очевидный алгоритм требует небольших вычислений, но срабатывает с пренебрежимо малой вероятностью (при больших длинах текстов). На самом деле противник может пытаться атаковать криптосистему различными способами и использовать различные, более изощрённые алгоритмы поиска открытого текста. Естественно считать криптосистему стойкой, если любой такой алгоритм требует практически неосуществимого объема вычислений или срабатывает с пренебрежимо малой вероятностью. (При этом противник может использовать не только детерминированные, но и вероятностные алгоритмы.) Это и есть теоретико-сложностной подход к определению стойкости. Для его реализации в отношении того или иного типа криптографических схем необходимо выполнить следующее:

1. Дать формальное определение схемы данного типа.
2. Дать формальное определение стойкости схемы.
3. Доказать стойкость конкретной конструкции схемы данного типа.

Здесь сразу же возникает ряд проблем.

Во-первых, в криптографических схемах, разумеется, всегда используются фиксированные значения параметров. Например, криптосистемы разрабатываются для ключей длины, скажем, в 256 или 512 байтов. Для применения же теоретико-сложностного подхода необходимо, чтобы задача, вычислительную сложность которой предполагается использовать, была массовой. Поэтому в теоретической криптографии рассматриваются математические модели криптографических схем. Эти модели зависят от некоторого параметра, называемого параметром безопасности, который может принимать сколь угодно большие значения (обычно для простоты предполагается, что параметр безопасности может пробегать весь натуральный ряд).

Во-вторых, определение стойкости криптографической схемы зависит от той задачи, которая стоит перед противником, и от того, какая информация о схеме ему доступна. Поэтому стойкость схем приходится определять и исследовать отдельно для каждого предположения о противнике.

В-третьих, необходимо уточнить, какой объем вычислений можно считать «практически неосуществимым». Из сказанного выше следует, что эта величина не может быть просто константой, она должна быть представлена функцией от растущего параметра безопасности. В соответствии с тезисом Эдмондса алгоритм считается эффективным, если время его выполнения ограничено некоторым полиномом от длины входного слова (в нашем случае — от параметра безопасности). В противном случае говорят, что вычисления по данному алгоритму практически неосуществимы. Заметим также, что сами криптографические схемы должны быть эффективными, т. е. все вычисления, предписанные той или иной схемой, должны выполняться за полиномиальное время.

В-четвёртых, необходимо определить, какую вероятность можно считать пренебрежимо малой. В криптографии принято считать таковой любую вероятность, которая для любого полинома p и для всех достаточно больших n не превосходит $1/p(n)$, где n — параметр безопасности.

Итак, при наличии всех указанных выше определений, проблема обоснования стойкости криптографической схемы свелась к доказательству отсутствия полиномиального алгоритма, который решает задачу, стоящую перед противником. Но здесь возникает ещё одно и весьма серьёзное препятствие: современное состояние теории сложности вычислений не позволяет доказывать сверхполиномиальные нижние оценки сложности для конкретных задач рассматриваемого класса. Из этого следует, что на данный момент стойкость криптографических схем может быть установлена лишь с привлечением каких-либо недоказанных предположений. Поэтому основное направление исследований состоит в поиске

наиболее слабых достаточных условий (в идеале — необходимых и достаточных) для существования стойких схем каждого из типов. В основном рассматриваются предположения двух типов — общие (или теоретико-сложностные) и теоретико-числовые, т. е. предположения о сложности конкретных теоретико-числовых задач. Все эти предположения в литературе обычно называются криптографическими.

Ниже мы кратко рассмотрим несколько интересных математических объектов, возникших на стыке теории сложности и криптографии. Более подробный обзор по этим вопросам можно найти в книге [1].

2. Криптография и гипотеза P≠NP

Как правило, знакомство математиков-неспециалистов с теорией сложности вычислений ограничивается классами P и NP и знаменитой гипотезой P≠NP.

Напомним вкратце необходимые сведения из теории сложности вычислений. Пусть Σ — множество всех конечных строк в двоичном алфавите. Подмножества $L \subseteq \Sigma$ в теории сложности принято называть языками. Говорят, что машина Тьюринга M работает за полиномиальное время (или просто, что она полиномиальна), если существует полином p такой, что на любом входном слове длины n машина M останавливается после выполнения не более, чем $p(n)$ операций. Машина Тьюринга M распознает (другой термин — принимает) язык L, если на всяком входном слове $x \in L$ машина M останавливается в принимающем состоянии, а на всяком слове $x \notin L$ — в отвергающем. Класс P — это класс всех языков, распознаваемых машинами Тьюринга, работающими за полиномиальное время. Функция $f : \Sigma \to \Sigma$ вычислима за полиномиальное время, если существует полиномиальная машина Тьюринга такая, что если на вход ей подано слово $x \in \Sigma$, то в момент останова на ленте будет записано значение $f(x)$. Язык L принадлежит классу NP, если существуют предикат $P(x, y) : \Sigma \times \Sigma \to \{0, 1\}$, вычислимый за полиномиальное время, и полином p такие, что $L = \{x | \exists y P(x, y) \& |y| \leqslant p(|x|)\}$. Таким образом, язык L принадлежит NP, если для всякого слова из L длины n можно угадать некоторую строку полиномиальной от n длины и затем с помощью предиката P убедиться в правильности догадки. Ясно, что P⊆NP. Является ли это включение строгим — одна из самых известных нерешённых задач математики. Большинство специалистов считают, что оно строгое (так называемая гипотеза P≠NP). В классе NP выделен подкласс максимально сложных языков, называемых NP-полными: любой NP-полный язык распознаваем за полиномиальное время тогда и только тогда, когда P=NP.

Для дальнейшего нам потребуется ещё понятие вероятностной машины Тьюринга. В обычных машинах Тьюринга (их называют детерминированными, чтобы отличить от вероятностных) новое состояние, в которое машина переходит на очередном шаге, полностью определяется текущим состоянием и тем символом, который обозревает головка на ленте. В вероятностных машинах новое состояние может зависеть ещё и от случайной величины, которая принимает

значения 0 и 1 с вероятностью 1/2 каждое. Альтернативно, можно считать, что вероятностная машина Тьюринга имеет дополнительную случайную ленту, на которой записана бесконечная двоичная случайная строка. Случайная лента может читаться только в одном направлении и переход в новое состояние может зависеть от символа, обозреваемого на этой ленте.

Рассмотрим теперь следующий естественный вопрос: не является ли гипотеза P≠NP необходимым и достаточным условием для существования стойких криптографических схем?

Необходимость, и в самом деле, во многих случаях почти очевидна. Вернемся к примеру 1. Определим следующий язык

$L = \{(K_1, d, i)|$ существует сообщение m такое, что $E_{K_1}(m) = d$ и его i-ый бит равен 1$\}$.

Ясно, что $L \in$ NP: вместо описанного во введении полного перебора можно просто угадать открытый текст m и проверить за полиномиальное время, что $E_{K_1}(m) = d$ и i-ый бит m равен 1. Если да, то входное слово (K_1, d, i) принимается, в противном случае — отвергается.

В предположении P=NP существует детерминированный полиномиальный алгоритм, распознающий язык L. Зная K_1 и d, с помощью этого алгоритма можно последовательно, по биту, вычислить открытый текст m. Тем самым криптосистема нестойкая.

Тот же подход: угадать секретный ключ и проверить (за полиномиальное время) правильность догадки, применим в принципе и к другим криптографическим схемам. Однако, в некоторых случаях возникают технические трудности, связанные с тем, что по информации, которая имеется у противника, искомая величина (открытый текст, секретный ключ и т. п.) восстанавливается неоднозначно.

Что же касается вопроса о достаточности предположения P≠NP, то здесь напрашивается следующий подход: выбрать какую-либо NP-полную задачу и построить на её основе криптографическую схему, задача взлома которой (т. е. задача, стоящая перед противником) была бы NP-полной. Такие попытки предпринимались в начале 80-х годов, в особенности в отношении криптосистем с открытым ключом, но к успеху не привели. Результатом всех этих попыток стало осознание следующего факта: даже если P≠NP, то любая NP-полная задача может оказаться трудной лишь на некоторой бесконечной последовательности входных слов. Иными словами, в определение класса NP заложена мера сложности «в худшем случае». Для стойкости же криптографической схемы необходимо, чтобы задача противника была сложной «почти всюду». Таким образом, стало ясно, что для криптографической стойкости необходимо существенно более сильное предположение, чем P≠NP. А именно, предположение о существовании односторонних функций.

3. Односторонние функции

Говоря неформально, односторонняя функция — это эффективно вычислимая функция, для задачи инвертирования которой не существует эффективных алгоритмов. Под инвертированием понимается массовая задача нахождения по заданному значению функции одного (любого) значения из прообраза (заметим, что обратная функция, вообще говоря, может не существовать).

Поскольку понятие односторонней функции — центральное в математической криптографии, ниже мы даем его формальное определение.

Пусть $\Sigma^n = \{0,1\}^n$ — множество всех двоичных строк длины n. Под функцией f мы понимаем семейство $\{f_n\}$, где $f_n : \Sigma^n \to \Sigma^m$, $m = m(n)$. Для простоты изложения мы предполагаем, что n пробегает весь натуральный ряд и что каждая из функций f_n всюду определена.

Функция f называется *честной*, если существует полином q такой, что $n \leqslant q(m(n))$ для всех n.

Определение 1. Честная функция f называется *односторонней*, если

1. Существует полиномиальный алгоритм, который для всякого x вычисляет $f(x)$.

2. Для любой полиномиальной вероятностной машины Тьюринга A выполнено следующее. Пусть строка x выбрана наудачу из множества Σ^n (обозначается $x \in_R \Sigma^n$). Тогда для любого полинома p и всех достаточно больших n

$$Pr\{f(A(f(x))) = f(x)\} < 1/p(n).$$

Вероятность здесь определяется случайным выбором строки x и случайными величинами, которые A использует в своей работе.

Условие 2 качественно означает следующее. Любая полиномиальная вероятностная машина Тьюринга A может по данному y найти x из уравнения $f(x) = y$ лишь с пренебрежимо малой вероятностью.

Заметим, что требование честности нельзя опустить. Поскольку длина входного слова $f(x)$ машины A равна m, ей может не хватить полиномиального (от m) времени просто на выписывание строки x, если f слишком сильно «сжимает» входные значения.

Ясно, что из предположения о существовании односторонних функций следует, что P≠NP. Однако, не исключена следующая ситуация: P≠NP, но односторонних функций нет.

Существование односторонних функций является необходимым условием для стойкости многих типов криптографических схем. В некоторых случаях этот факт устанавливается достаточно просто. Обратимся опять

к примеру 1. Рассмотрим функцию f такую, что $f(r) = K_1$. Она вычислима с помощью алгоритма G за полиномиальное время. Покажем, что если f — не односторонняя функция, то криптосистема нестойкая. Предположим, что существует полиномиальный вероятностный алгоритм A, который инвертирует f с вероятностью по крайней мере $1/p(n)$ для некоторого полинома p. Здесь n — длина ключа K_1. Противник может подать на вход алгоритму A ключ K_1 и получить с указанной вероятностью некоторое значение r' из прообраза. Далее противник подаёт r' на вход алгоритма G и получает пару ключей (K_1, K_2'). Хотя K_2' не обязательно совпадает с K_2, тем не менее по определению криптосистемы $D_{K_2'}(E_{K_1}(m)) = m$ для любого открытого текста m. Поскольку K_2' найден с вероятностью по крайней мере $1/p(n)$, которая в криптографии не считается пренебрежимо малой, криптосистема нестойкая.

Для других криптографических схем подобный результат доказывается не столь просто. В работе Импальяццо и Луби [7] доказана необходимость односторонних функций для существования целого ряда стойких криптографических схем.

Из всего сказанного следует, что предположение о существовании односторонних функций является самым слабым криптографическим предположением, которое может оказаться достаточным для доказательства существования стойких криптографических схем различных типов. На выяснение того, является ли это условие и в самом деле достаточным, направлены значительные усилия специалистов. Трудность задачи построения криптографических схем из односторонних функций можно пояснить на следующем примере. Пусть f — односторонняя функция и нам требуется построить *криптосистему с секретным ключом*. В такой криптосистеме имеется только один ключ — секретный, который известен и отправителю, и получателю шифрованного сообщения. Алгоритмы шифрования E_K и дешифрования D_K оба зависят от этого секретного ключа K и таковы, что $D_K(E_K(m)) = m$ для любого открытого текста m. Ясно, что если криптограмму d сообщения m вычислять как $d = f(m)$, то противник, перехвативший d, может вычислить m лишь с пренебрежимо малой вероятностью. Но во-первых, непонятно, каким образом сможет восстановить сообщение m из криптограммы законный получатель? Во-вторых, из того, что функция f односторонняя следует лишь, что противник не может вычислить всё сообщение целиком. А это — весьма низкий уровень стойкости. Желательно, чтобы противник, знающий криптограмму d, не мог вычислить ни одного бита открытого текста.

На настоящий момент доказано, что существование односторонних функций является необходимым и достаточным условием для существования стойких криптосистем с секретным ключом, а также стойких

криптографических протоколов нескольких типов, включая протоколы электронной подписи. С другой стороны, имеется результат Импальяццо и Рудиха [9], который является достаточно сильным аргументом в пользу того, что для некоторых типов криптографических схем (включая протоколы распределения ключей типа Диффи-Хеллмана) требуются более сильные предположения, чем предположение о существовании односторонних функций. К сожалению, этот результат слишком сложный, чтобы его можно было разъяснить в настоящей статье.

4. Псевдослучайные генераторы

Существенный недостаток шифра Вернама состоит в том, что ключи одноразовые. Можно ли избавиться от этого недостатка за счёт некоторого снижения стойкости? Один из способов решения этой проблемы состоит в следующем. Отправитель и получатель имеют общий секретный ключ K длины n и с помощью некоторого достаточно эффективного алгоритма g генерируют из него последовательность $r = g(K)$ длины $q(n)$, где q — некоторый полином. Такая криптосистема (обозначим её Cr) позволяет шифровать сообщение m (или совокупность сообщений) длиной до $q(n)$ битов по формуле $d = r \oplus m$, где $\oplus$ — поразрядное сложение битовых строк по модулю 2. Дешифрование выполняется по формуле $m = d \oplus r$. Из результатов Шеннона вытекает, что такая криптосистема не является абсолютно стойкой, т. е. стойкой против любого противника (в чем, впрочем, нетрудно убедиться и непосредственно). Но что будет, если требуется защищаться только от полиномиально ограниченного противника, который может атаковать криптосистему лишь с помощью полиномиальных вероятностных алгоритмов? Каким условиям должны удовлетворять последовательность r и алгоритм g, чтобы криптосистема Cr была стойкой? Поиски ответов на эти вопросы привели к появлению понятия псевдослучайного генератора, которое было введено Блюмом и Микали [3].

Пусть $g : \{0,1\}^n \to \{0,1\}^{q(n)}$ — функция, вычислимая за полиномиальное (от n) время. Такая функция называется генератором. Интуитивно, генератор g является псевдослучайным, если порождаемые им последовательности неотличимы никаким полиномиальным вероятностным алгоритмом от случайных последовательностей той же длины $q(n)$. Формально этот объект определяется следующим образом.

Пусть A — полиномиальная вероятностная машина Тьюринга, которая получает на входе двоичные строки длины $q(n)$ и выдаёт в результате своей работы один бит. Пусть

$$P_1(A, n) = Pr\{A(r) = 1 | r \in_R \{0,1\}^{q(n)}\}.$$

Вероятность здесь определяется случайным выбором строки r и случайными величинами, которые A использует в своей работе. Пусть

$$P_2(A,n) = Pr\{A(g(s)) = 1 | s \in_R \{0,1\}^n\}.$$

Эта вероятность определяется случайным выбором строки s и случайными величинами, которые A использует в своей работе. Подчеркнём, что функция g вычисляется детерминированным алгоритмом.

Определение 2. Генератор g называется *криптографически стойким псевдослучайным генератором*, если для любой полиномиальной вероятностной машины Тьюринга A, для любого полинома p и всех достаточно больших n

$$|P_1(A,n) - P_2(A,n)| < 1/p(n).$$

Всюду ниже мы для краткости будем называть криптографически стойкие псевдослучайные генераторы просто псевдослучайными генераторами. Такое сокращение является общепринятым в криптографической литературе.

Нетрудно убедиться, что для существования псевдослучайных генераторов необходимо существование односторонних функций. В самом деле, сама функция g должна быть односторонней. Доказательство этого простого факта мы оставляем читателю в качестве упражнения. Вопрос о том, является ли существование односторонних функций одновременно и достаточным условием, долгое время оставался открытым. В 1982 г. Яо [10] построил псевдослучайный генератор, исходя из предположения о существовании *односторонних перестановок*, т. е. сохраняющих длину взаимнооднозначных односторонних функций. За этим последовала серия работ, в которых достаточное условие всё более и более ослаблялось, пока наконец в 1989–1990 гг. Импальяццо, Левин и Луби [8] и Хостад [6] не получили следующий окончательный результат.

Теорема 1. *Псевдослучайные генераторы существуют тогда и только тогда, когда существуют односторонние функции.*

Псевдослучайные генераторы находят применение не только в криптографии, но и в теории сложности, и в других областях дискретной математики. Обсуждение этих приложений выходит за рамки настоящей статьи. Здесь же в качестве иллюстрации мы рассмотрим описанную в начале данного раздела криптосистему Cr, использующую псевдослучайный генератор в качестве алгоритма g. Прежде всего, нам необходимо дать определение стойкости криптосистемы с секретным ключом.

Пусть E_K — алгоритм шифрования криптосистемы с секретным ключом. Обозначим результат его работы $d = E_K(m)$, здесь K — секретный

ключ длиной n битов, а m — открытый текст длиной $q(n)$ битов. Через m_i обозначается i-ый бит открытого текста. Пусть A — полиномиальная вероятностная машина Тьюринга, которая получает на вход криптограмму d и выдаёт пару (i, σ), где $i \in \{1, \ldots, q(n)\}$, $\sigma \in \{0, 1\}$. Интуитивно, криптосистема является стойкой, если никакая машина Тьюринга A не может вычислить ни один бит открытого текста с вероятностью успеха, существенно большей, чем при простом угадывании.

Определение 3. Криптосистема называется стойкой, если для любой полиномиальной вероятностной машины Тьюринга A, для любого полинома p и всех достаточно больших n

$$Pr\{A(c) = (i, \sigma) \ \&\ \sigma = m_i \,\big|\, K \in_R \{0,1\}^n,\ m \in_R \{0,1\}^{q(n)}\} < 1/2 + 1/p(n).$$

Эта вероятность (всюду ниже для краткости мы её обозначаем просто Pr) определяется случайным выбором секретного ключа K, случайным выбором открытого текста m из множества всех двоичных строк длины $q(n)$ и случайными величинами, которые A использует в своей работе.

Покажем, что криптосистема Cr с псевдослучайным генератором в качестве g является стойкой в смысле данного определения. Предположим противное, т. е. что существуют полиномиальный вероятностный алгоритм A и полином p такие, что $Pr \geqslant 1/2 + 1/p(n)$ для бесконечно многих n. Рассмотрим алгоритм B, который получает на входе двоичную строку r длины $q(n)$, выбирает $m \in_R \{0,1\}^{q(n)}$, вычисляет $d = m \oplus r$ и вызывает A как подпрограмму, подавая ей на вход строку d. Получив от A пару (i, σ), B проверяет, действительно ли $m_i = \sigma$ и если да, то выдаёт 1, в противном случае — 0, и останавливается. Легко видеть, что B работает за полиномиальное (от n) время. Убедимся, что алгоритм B отличает псевдослучайные строки, порождённые генератором g, от случайных строк длины $q(n)$. В самом деле, если строки r, поступающие на вход B, являются случайными, то d — криптограмма шифра Вернама и, согласно теореме Шеннона, $Pr = 1/2$. Если строки r порождены генератором g, то криптограммы d имеют такое же распределение вероятностей, как в криптосистеме Cr, и, согласно предположению, $Pr \geqslant 1/2 + 1/p(n)$ для бесконечно многих n. Полученное противоречие с определением псевдослучайного генератора доказывает утверждение о стойкости криптосистемы Cr.

Разумеется, стойкость криптосистемы с секретным ключом можно определять различным образом. Например, можно рассматривать стойкость против атаки с выбором открытого текста: противник может предварительно выбрать полиномиальное количество открытых текстов и получить их криптограммы, после чего он получает ту криптограмму, по

которой ему требуется вычислить хотя бы один бит соответствующего открытого текста. Нетрудно убедиться, что криптосистема Cr с псевдослучайным генератором в качестве g является стойкой и против атаки с выбором открытого текста.

Таким образом, мы убедились, что с помощью псевдослучайных генераторов можно строить стойкие криптосистемы. Основное направление исследований в данной области — поиск методов построения эффективных псевдослучайных генераторов на основе различных криптографических предположений. Показателем эффективности здесь служит количество операций, затрачиваемых на вычисление каждого очередного бита псевдослучайной последовательности.

5. Доказательства с нулевым разглашением

Предположим, что Алиса знает доказательство некоторой теоремы и желает убедить Боба в том, что теорема верна. Конечно, Алиса может просто передать доказательство Бобу на проверку. Но тогда впоследствии Боб сможет сам, без помощи Алисы, доказывать третьим лицам эту теорему. А может ли Алиса убедить Боба так, чтобы он не получил при этом никакой информации, которая помогла бы ему восстановить доказательство теоремы? Этим двум, казалось бы взаимно исключающим требованиям, удовлетворяют протоколы доказательства с нулевым разглашением. Последнее понятие было введено Гольдвассер, Микали и Ракоффом в 1985 г. [4].

Рассматривается следующая модель протокола. В распоряжении Алисы и Боба имеются вероятностные машины Тьюринга $\mathbf{P}$ и $\mathbf{V}$ соответственно. Вычислительные ресурсы, которые может использовать Алиса, неограничены, в то время как машина $\mathbf{V}$ работает за полиномиальное время. Машины $\mathbf{P}$ и $\mathbf{V}$ имеют общую коммуникационную ленту для обмена сообщениями. После записи сообщения на коммуникационную ленту машина переходит в состояние ожидания и выходит из него, как только на ленту будет записано ответное сообщение. Машины $\mathbf{P}$ и $\mathbf{V}$ имеют также общую входную ленту, на которую записано входное слово x. Утверждение, которое доказывает Алиса, суть «$x \in L$», где L — некоторый фиксированный (известный и Алисе, и Бобу) язык. Чтобы избежать тривиальности, язык L должен быть трудным (например, NP-полным), иначе Боб сможет самостоятельно проверить, что $x \in L$. По существу, протокол доказательства состоит в том, что Боб, используя случайность, выбирает некоторые вопросы, задаёт их Алисе и проверяет правильность ответов. Выполнение протокола завершается, когда машина $\mathbf{V}$ останавливается,

при этом она выдаёт 1, если доказательство принято, и 0 — в противном случае.

Пусть A и B — две интерактивные, т. е. взаимодействующие через общую коммуникационную ленту, вероятностные машины Тьюринга. Через $[B(x), A(x)]$ обозначается случайная величина — выходное слово машины A, когда A и B работают на входном слове x. Через $|x|$ обозначается длина слова x.

Определение 4. *Интерактивным доказательством для языка* L называется пара интерактивных машин Тьюринга $(\mathbf{P}, \mathbf{V})$ такая, что выполняются следующие два условия.

1. (Полнота). Для всех $x \in L$

$$Pr\{[\mathbf{P}(x), \mathbf{V}(x)] = 1\} = 1.$$

2. (Корректность). Для любой машины Тьюринга $\mathbf{P}^*$, для любого полинома p и для всех $x \notin L$ достаточно большой длины

$$Pr\{[\mathbf{P}^*(x), \mathbf{V}(x)] = 1\} < 1/p(|x|).$$

Полнота означает, что если входное слово принадлежит языку L и оба участника, и Алиса, и Боб, следуют протоколу, то доказательство будет всегда принято. Требование корректности защищает Боба от нечестной Алисы, которая пытается обмануть его, «доказывая» ложное утверждение. При этом Алиса может каким угодно образом отклоняться от действий, предписанных протоколом, т. е. вместо машины Тьюринга $\mathbf{P}$ использовать любую другую машину $\mathbf{P}^*$. Требуется, чтобы вероятность обмана была в любом случае пренебрежимо малой.

Определение 5. Интерактивный протокол доказательства для языка L называется *доказательством с абсолютно нулевым разглашением*, если, кроме условий 1 и 2, выполнено ещё и следующее условие.

3. (Свойство нулевого разглашения). Для любой полиномиальной вероятностной машины Тьюринга $\mathbf{V}^*$ существует вероятностная машина Тьюринга $\mathbf{M}_{\mathbf{V}^*}$, работающая за полиномиальное в среднем время, и такая, что для всех $x \in L$

$$\mathbf{M}_{\mathbf{V}^*}(x) = [\mathbf{P}(x), \mathbf{V}^*(x)].$$

Машина $\mathbf{M}_{\mathbf{V}^*}$ называется моделирующей машиной для $\mathbf{V}^*$. Предполагается, что математическое ожидание времени её работы ограничено полиномом от длины x. Это означает, что в принципе $\mathbf{M}_{\mathbf{V}^*}$ может, в

зависимости от того, какие значения примут используемые в её работе случайные переменные, работать достаточно долго. Но вероятность того, что время её работы превысит некоторую полиномиальную границу, мала. Для каждой машины $\mathbf{V}^*$ строится своя моделирующая машина; последняя может использовать $\mathbf{V}^*$ как подпрограмму. Через $\mathbf{M}_{\mathbf{V}^*}(x)$ обозначается случайная величина — выходное слово машины $\mathbf{M}_{\mathbf{V}^*}$, когда на входе она получает слово x.

Свойство нулевого разглашения защищает Алису от нечестного Боба, который, произвольно отклоняясь от действий, предписанных протоколом (используя $\mathbf{V}^*$ вместо $\mathbf{V}$), пытается извлечь из его выполнения дополнительную информацию. Условие 3 означает, что Боб может при этом получить только такую информацию, которую он смог бы вычислить и самостоятельно (без выполнения протокола) за полиномиальное время.

Приведем в качестве примера протокол доказательства с абсолютно нулевым разглашением для языка ИЗОМОРФИЗМ ГРАФОВ из работы Гольдрайха, Микали и Вигдерсона [5]. Входным словом является пара графов $G_1 = (U, E_1)$ и $G_0 = (U, E_0)$. Здесь U — множество вершин, которое можно отождествить с множеством натуральных чисел $\{1, \ldots, n\}$, E_1 и E_0 — множества рёбер такие, что $|E_1| = |E_0| = m$. Графы G_1 и G_0 называются изоморфными, если существует перестановка φ на множестве U такая, что $(u, v) \in E_0$ тогда и только тогда, когда $(\varphi(u), \varphi(v)) \in E_1$ (обозначается $G_1 = \varphi G_0$). Задача распознавания изоморфизма графов — хорошо известная математическая задача, для которой на данный момент не известно полиномиальных алгоритмов. С другой стороны, неизвестно, является ли эта задача NP-полной, хотя есть веские основания предполагать, что не является.

Протокол IG

Пусть φ — изоморфизм между G_1 и G_0. Следующие четыре шага выполняются в цикле m раз, каждый раз с независимыми случайными величинами.

1. $\mathbf{P}$ *выбирает случайную перестановку π на множестве U, вычисляет $H = \pi G_1$ и посылает этот граф* $\mathbf{V}$.

2. $\mathbf{V}$ *выбирает случайный бит α и посылает его* $\mathbf{P}$.

3. Если $\alpha = 1$, то $\mathbf{P}$ *посылает* $\mathbf{V}$ *перестановку π, в противном случае — перестановку $\pi \circ \varphi$.*

4. Если перестановка, полученная $\mathbf{V}$, *не является изоморфизмом между G_α и H, то* $\mathbf{V}$ *останавливается и отвергает доказательство. В противном случае выполнение протокола продолжается.*

Если проверки п.4 дали положительный результат во всех m циклах, то $\mathbf{V}$ *принимает доказательство.*

Заметим, что если в протоколе IG машина $\mathbf{P}$ получает изоморфизм φ в качестве дополнительного входного слова, то ей для выполнения протокола не требуются неограниченные вычислительные ресурсы. Более того, в этом случае $\mathbf{P}$ может быть полиномиальной вероятностной машиной Тьюринга.

Теорема 2 ([5]). *Протокол IG является доказательством с абсолютно нулевым разглашением для языка ИЗОМОРФИЗМ ГРАФОВ.*

Полнота протокола IG очевидна.

Для доказательства корректности достаточно заметить, что бит α, который $\mathbf{V}$ выбирает на шаге 2, указывает $\mathbf{P}$, для какого из графов — G_0 или G_1 — требуется продемонстрировать изоморфизм с графом H. Если G_0 и G_1 не изоморфны, то H может быть изоморфен, в лучшем случае, одному из них. Поэтому проверка п. 4 даст положительный результат с вероятностью $\leqslant 1/2$ в одном цикле и с вероятностью $\leqslant 1/2^m$ во всех m циклах.

Доказательство свойства нулевого разглашения значительно сложнее. Поэтому мы воспроизводим только основную идею. Прежде всего, заметим, что основная задача машины $\mathbf{V}^*$ — получить максимально возможную информацию об изоморфизме между G_0 и G_1. Естественно предположить, что она, в отличие от $\mathbf{V}$, будет выдавать в качестве выходного слова не один бит, а всю полученную в результате выполнения протокола информацию, включая графы H и перестановки, полученные соответственно на шагах 1 и 3 протокола IG. Моделирующая машина $\mathbf{M}_{\mathbf{V}^*}$ должна уметь строить такие же графы и перестановки, не зная при этом изоморфизм φ! Поэтому $\mathbf{M}_{\mathbf{V}^*}$ пытается угадать тот бит α, который будет запросом машины $\mathbf{V}^*$ на шаге 2. Для этого $\mathbf{M}_{\mathbf{V}^*}$ выбирает случайный бит β, случайную перестановку ψ и вычисляет $H = \psi G_\beta$. Далее $\mathbf{M}_{\mathbf{V}^*}$ запоминает состояние машины $\mathbf{V}^*$ и вызывает её как подпрограмму, подавая ей на вход граф H. Ответом машины $\mathbf{V}^*$ будет некоторый бит α. Если $\alpha = \beta$, то моделирование в данном цикле завершено успешно, поскольку $\mathbf{M}_{\mathbf{V}^*}$ может продемонстрировать требуемый изоморфизм. Если же $\alpha \neq \beta$, то $\mathbf{M}_{\mathbf{V}^*}$ восстанавливает ранее сохранённое состояние машины $\mathbf{V}^*$ и повторяет попытку.

Если в определении свойства нулевого разглашения заменить равенство случайных величин $\mathbf{M}_{\mathbf{V}^*}(x)$ и $[\mathbf{P}(x), \mathbf{V}^*(x)]$ требованием, чтобы их распределения вероятностей «почти не отличались», то получится дру-

гая разновидность доказательств — *доказательства со статистически нулевым разглашением.*

Еще один тип — *доказательства с вычислительно нулевым разглашением.* В этом случае требуется, чтобы моделирующая машина создавала распределение вероятностей, которое неотличимо от $[\mathbf{P}(x), \mathbf{V}^*(x)]$ никаким полиномиальным вероятностным алгоритмом (неотличимость здесь определяется аналогично тому, как это делалось в определении псевдослучайного генератора).

Подчеркнём особо, что во всех трёх определениях нулевого разглашения условия накладываются на действия моделирующей машины только на тех словах, которые принадлежат языку.

Помимо интереса к доказательствам с нулевым разглашением как к нетривиальному математическому объекту, они исследуются также и в связи с практическими приложениями. Наиболее естественный и важный тип таких приложений — протоколы аутентификации. С помощью такого протокола Алиса может доказать Бобу свою аутентичность. Предположим, например, что Алиса — это интеллектуальная банковская карточка, в которой реализован алгоритм $\mathbf{P}$, а Боб — это компьютер банка, выполняющий программу $\mathbf{V}$. Прежде чем начать выполнение каких-либо банковских операций, банк должен убедиться в подлинности карточки и идентифицировать её владельца, или, говоря на языке криптографии, карточка должна пройти аутентификацию. В принципе для этой цели можно использовать приведенный выше протокол IG. В этом случае в памяти банковского компьютера хранится пара графов (G_0, G_1), сопоставленная Алисе, а на интеллектуальной карточке — та же пара графов и изоморфизм φ. Предполагается, что, кроме Алисы, этот изоморфизм никто не знает (кроме, быть может, Боба) и поэтому с помощью протокола IG карточка доказывает свою аутентичность. При этом свойство полноты означает, что карточка наверняка докажет свою аутентичность. Свойство корректности защищает интересы банка от злоумышленника, который, не являясь клиентом банка, пытается пройти аутентификацию, используя фальшивую карточку. Свойство нулевого разглашения защищает клиента от злоумышленника, который, подслушав одно или более выполнений протокола аутентификации данной карточки, пытается пройти аутентификацию под именем Алисы. Для практического применения очень важным свойством протокола IG является то, что алгоритм $\mathbf{P}$, получивший в качестве дополнительного входа изоморфизм φ, работает за полиномиальное время. Вместо протокола IG можно использовать, вообще говоря, любое другое доказательство с нулевым разглашением, в котором алгоритм $\mathbf{P}$ обладает этим свойством. Но для реальных приложений протокол IG, как и большинство подобных протоколов, не

эффективен: большое количество циклов, слишком длинные сообщения и т. д. Поиск более эффективных доказуемо стойких протоколов — одно из основных направлений исследований в данной области.

Список литературы

[1] *Анохин М. И., Варновский Н. П., Сидельников В. М., Ященко В. В.* Криптография в банковском деле. М.: МИФИ, 1997.

[2] *Гэри М., Джонсон Д.* Вычислительные машины и трудно решаемые задачи. М.: Мир, 1982.

[3] *Blum M., Micali S.* How to generate crytographically strong sequences of pseudo-random bits // SIAM J. Comput. **V. 13**, No 4, 1984. P. 850–864.

[4] *Goldwasser S., Micali S., Rackoff C.* The knowledge complexity of interactive proof systems // SIAM J. Comput. **V. 18**, No 1, 1989. P. 186–208.

[5] *Goldreich O., Micali S., Wigderson A.* Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems // J. ACM. **V. 38**, No 3, 1991. P. 691–729.

[6] *Håstad J.* Pseudo-random generators under uniform assumptions // Proc. 22nd Annu. ACM Symp. on Theory of Computing. 1990. P. 395–404.

[7] *Impagliazzo R., Luby M.* One-way functions are essential for complexity based cryptography // Proc. 30th Annu. Symp. on Found. of Comput. Sci. 1989. P. 230–235.

[8] *Impagliazzo R. , Levin L., Luby M.* Pseudo-random generation from one-way functions // Proc. 21st Annu. ACM Symp. on Theory of Computing. 1989. P. 12–24.

[9] *Impagliazzo R., Rudich S.* Limits on the provable consequences of one-way permutations // Proc. 21st Annu. ACM Symp. on Theory of Computing. 1989. P. 44–61.

[10] *Yao A.C.* Theory and applications of trapdoor functions // Proc. 23rd Annu. Symp. on Found. of Comput. Sci. 1982. P. 80–91.

Алгоритмические проблемы теории чисел

Ю. В. Нестеренко

Эта статья посвящена алгоритмам теории чисел. Вопрос «как сосчитать?» всегда сопутствовал теоретико-числовым исследованиям. Труды Евклида и Диофанта, Ферма и Эйлера, Гаусса, Чебышева и Эрмита содержат остроумные и весьма эффективные алгоритмы решения диофантовых уравнений, выяснения разрешимости сравнений, построения больших по тем временам простых чисел, нахождения наилучших приближений и т.д. Без преувеличения можно сказать, что вся теория чисел пронизана алгоритмами. В последние два десятилетия, благодаря в первую очередь запросам криптографии и широкому распространению ЭВМ, исследования по алгоритмическим вопросам теории чисел переживают период бурного и весьма плодотворного развития. Мы кратко затронем здесь лишь те алгоритмические аспекты теории чисел, которые связаны с криптографическими применениями. За рамками статьи останутся проблемы нахождения решений диофантовых уравнений, вычислений в полях алгебраических чисел, вычислений с решётками, нахождения диофантовых приближений и ряд других вопросов.

Вычислительные машины и электронные средства связи проникли практически во все сферы человеческой деятельности. Немыслима без них и современная криптография. Шифрование и дешифрование текстов можно представлять себе как процессы переработки целых чисел при помощи ЭВМ, а способы, которыми выполняются эти операции, как некоторые функции, определённые на множестве целых чисел. Всё это делает естественным появление в криптографии методов теории чисел. Кроме того, стойкость ряда современных криптосистем обосновывается только сложностью некоторых теоретико-числовых задач (см. [24]).

Но возможности ЭВМ имеют определённые границы. Приходится разбивать длинную цифровую последовательность на блоки ограниченной длины и шифровать каждый такой блок отдельно. Мы будем считать в дальнейшем, что все шифруемые целые числа неотрицательны и по величине меньше некоторого заданного (скажем, техническими ограничениями) числа m. Таким же условиям будут удовлетворять и числа, получаемые в процессе шифрования. Это позволяет считать и те, и другие числа

элементами кольца вычетов $\mathbb{Z}/m\mathbb{Z}$. Шифрующая функция при этом может рассматриваться как взаимнооднозначное отображение колец вычетов

$$f : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z},$$

а число $f(x)$ представляет собой сообщение x в зашифрованном виде.

Простейший шифр такого рода — шифр замены (см. [1]), соответствует отображению $f : x \to x + k \pmod m$ при некотором фиксированном целом k. Подобный шифр использовал ещё Юлий Цезарь. Конечно, не каждое отображение f подходит для целей надежного сокрытия информации, см. [1].

В 1978 г., см. [2], американцы Р. Ривест, А. Шамир и Л. Адлеман (R.L.Rivest, A.Shamir, L.Adleman) предложили пример функции f, обладающей рядом замечательных достоинств. На её основе была построена реально используемая система шифрования, получившая название по первым буквам имен авторов — система RSA. Эта функция такова, что

а) существует достаточно быстрый алгоритм вычисления значений $f(x)$;

б) существует достаточно быстрый алгоритм вычисления значений обратной функции $f^{-1}(x)$;

в) функция $f(x)$ обладает некоторым «секретом», знание которого позволяет быстро вычислять значения $f^{-1}(x)$; в противном же случае вычисление $f^{-1}(x)$ становится трудно разрешимой в вычислительном отношении задачей, требующей для своего решения столь много времени, что по его прошествии зашифрованная информация перестает представлять интерес для лиц, использующих отображение f в качестве шифра.

Подробнее об отображениях такого сорта и возможностях их использования в криптографии рассказано в [1,9].

Еще до выхода из печати статьи [2] копия доклада в Массачусетском Технологическом институте, посвящённого системе RSA, была послана известному популяризатору математики М. Гарднеру, который в 1977 г. в журнале Scientific American опубликовал статью [3], посвящённую этой системе шифрования. В русском переводе заглавие статьи Гарднера звучит так: *Новый вид шифра, на расшифровку которого потребуются миллионы лет*. Именно статья [3] сыграла важнейшую роль в распространении информации об RSA, привлекла к криптографии внимание широких кругов неспециалистов и фактически способствовала бурному прогрессу этой области, произошедшему в последовавшие 20 лет.

1. Система шифрования RSA

В дальнейшем мы будем предполагать, что читатель знаком с элементарными фактами теории чисел. Тех же, кто хотел бы ознакомиться с ними или напомнить себе эти факты, мы отсылаем к книге [4].

Пусть m и e натуральные числа. Функция f, реализующая схему RSA, устроена следующим образом

$$f : x \to x^e \pmod{m}. \tag{1}$$

Для расшифровки сообщения $a = f(x)$ достаточно решить сравнение

$$x^e \equiv a \pmod{m}. \tag{2}$$

При некоторых условиях на m и e это сравнение имеет единственное решение x.

Для того, чтобы описать эти условия и объяснить, как можно найти решение, нам потребуется одна теоретико-числовая функция, так называемая функция Эйлера. Эта функция натурального аргумента m обозначается $\varphi(m)$ и равняется количеству целых чисел на отрезке от 1 до m, взаимно простых с m. Так $\varphi(1) = 1$ и $\varphi(p^r) = p^{r-1}(p-1)$ для любого простого числа p и натурального r. Кроме того, $\varphi(ab) = \varphi(a)\varphi(b)$ для любых натуральных взаимно простых a и b. Эти свойства позволяют легко вычислить значение $\varphi(m)$, если известно разложение числа m на простые сомножители.

Если показатель степени e в сравнении (2) взаимно прост с $\varphi(m)$, то сравнение (2) имеет единственное решение. Для того, чтобы найти его, определим целое число d, удовлетворяющее условиям

$$de \equiv 1 \pmod{\varphi(m)}, \quad 1 \leqslant d < \varphi(m). \tag{3}$$

Такое число существует, поскольку $(e, \varphi(m)) = 1$, и притом единственно. Здесь и далее символом (a, b) будет обозначаться наибольший общий делитель чисел a и b. Классическая теорема Эйлера, см. [4], утверждает, что для каждого числа x, взаимно простого с m, выполняется сравнение $x^{\varphi(m)} \equiv 1 \pmod{m}$ и, следовательно,

$$a^d \equiv x^{de} \equiv x \pmod{m}. \tag{4}$$

Таким образом, в предположении $(a, m) = 1$, единственное решение сравнения (2) может быть найдено в виде

$$x \equiv a^d \pmod{m}. \tag{5}$$

Если дополнительно предположить, что число m состоит из различных простых сомножителей, то сравнение (5) будет выполняться и без предположения $(a, m) = 1$. Действительно, обозначим $r = (a, m)$ и $s = m/r$. Тогда

$\varphi(m)$ делится на $\varphi(s)$, а из (2) следует, что $(x, s) = 1$. Подобно (4), теперь легко находим $x \equiv a^d \pmod{s}$. А кроме того, имеем $x \equiv 0 \equiv a^d \pmod{r}$. Получившиеся сравнения в силу $(r, s) = 1$ дают нам (5).

Функция (1), принятая в системе RSA, может быть вычислена достаточно быстро. Как это сделать, мы обсудим чуть ниже. Пока отметим лишь, что обратная к $f(x)$ функция $f^{-1} : x \to x^d \pmod{m}$ вычисляется по тем же правилам, что и $f(x)$, лишь с заменой показателя степени e на d. Таким образом, для функции (1) будут выполнены указанные выше свойства а) и б).

Для вычисления функции (1) достаточно знать лишь числа e и m. Именно они составляют открытый ключ для шифрования. А вот для вычисления обратной функции требуется знать число d, оно и является «секретом», о котором речь идёт в пункте в). Казалось бы, ничего не стоит, зная число m, разложить его на простые сомножители, вычислить затем с помощью известных правил значение $\varphi(m)$ и, наконец, с помощью (3) определить нужное число d. Все шаги этого вычисления могут быть реализованы достаточно быстро, за исключением первого. Именно разложение числа m на простые множители и составляет наиболее трудоемкую часть вычислений. В теории чисел несмотря на многолетнюю её историю и на очень интенсивные поиски в течение последних 20 лет, эффективный алгоритм разложения натуральных чисел на множители так и не найден. Конечно, можно, перебирая все простые числа до $\sqrt{m}$, и, деля на них m, найти требуемое разложение. Но, учитывая, что количество простых в этом промежутке, асимптотически равно $2\sqrt{m} \cdot (\ln m)^{-1}$, см. [5], гл. 5, находим, что при m, записываемом 100 десятичными цифрами, найдётся не менее $4 \cdot 10^{42}$ простых чисел, на которые придётся делить m при разложении его на множители. Очень грубые прикидки показывают, что компьютеру, выполняющему миллион делений в секунду, для разложения числа $m > 10^{99}$ таким способом на простые сомножители потребуется не менее, чем 10^{35} лет. Известны и более эффективные способы разложения целых чисел на множители, чем простой перебор простых делителей, но и они работают очень медленно. Таким образом, название статьи М. Гарднера вполне оправдано.

Авторы схемы RSA предложили выбирать число m в виде произведения двух простых множителей p и q, примерно одинаковых по величине. Так как

$$\varphi(m) = \varphi(pq) = (p-1)(q-1), \tag{6}$$

то единственное условие на выбор показателя степени e в отображении

(1) есть

$$(e, p-1) = (e, q-1) = 1. \tag{7}$$

Итак, лицо, заинтересованное в организации шифрованной переписки с помощью схемы RSA, выбирает два достаточно больших простых числа p и q. Перемножая их, оно находит число $m = pq$. Затем выбирается число e, удовлетворяющее условиям (7), вычисляется с помощью (6) число $\varphi(m)$ и с помощью (3) — число d. Числа m и e публикуются, число d остается секретным. Теперь любой может отправлять зашифрованные с помощью (1) сообщения организатору этой системы, а организатор легко сможет расшифровывать их с помощью (5).

Для иллюстрации своего метода Ривест, Шамир и Адлеман зашифровали таким способом некоторую английскую фразу. Сначала она стандартным образом (a=01, b=02, ..., z=26, пробел=00) была записана в виде целого числа x, а затем зашифрована с помощью отображения (1) при

$$m = 11438162575788886766932577997614661201021829672124236256256184293570$$
$$69352457338978305971235639587050589907514759929002687954354 1$$

и $e = 9007$. Эти два числа были опубликованы, причем дополнительно сообщалось, что $m = pq$, где p и q — простые числа, записываемые соответственно 64 и 65 десятичными знаками. Первому, кто расшифрует соответствующее сообщение

$$f(x) = 96869613754622061477140922254355882905759991124574319874695120930 8$$
$$16298225145708356931476622883989628013391990551829945157815154,$$

была обещана награда в 100\$.

Эта история завершилась спустя 17 лет в 1994 г., см. [6], когда D. Atkins, M. Graff, A. K. Lenstra и P. C. Leyland сообщили о расшифровке фразы, предложенной в [2]. Она[1)] была вынесена в заголовок статьи [6], а соответствующие числа p и q оказались равными

$$p = 3490529510847650949147849619903898133417764638493387843990820577,$$
$$q = 3276913299326670954996198819083446141317764296799294253979828853 3.$$

Интересующиеся могут найти детали вычислений в работе [6]. Здесь же мы отметим, что этот замечательный результат (разложение на множители 129-значного десятичного числа) был достигнут благодаря использованию алгоритма разложения чисел на множители, называемого

[1)] ***The magic words are squeamish ossifrage.*** Приведём перевод двух последних слов, входящих в эту, по всей видимости, бессмысленную фразу:
squeamish — брезгливый, привередливый, обидчивый;
ossifrage — скопа (вид птицы типа выпи).

методом квадратичного решета. Выполнение вычислений потребовало колоссальных ресурсов. В работе, возглавлявшейся четырьмя авторами проекта, и продолжавшейся после предварительной теоретической подготовки примерно 220 дней, на добровольных началах участвовало около 600 человек и примерно 1600 компьютеров, объединённых сетью Internet. Наконец, отметим, что премия в 100$ была передана в Free Software Foundation.

Описанная выше схема RSA ставит ряд вопросов, которые мы и попробуем обсудить ниже. Например, как проводить вычисления с большими числами, ведь стандартное математическое обеспечение не позволяет перемножать числа размером по 65 десятичных знаков? Как вычислять огромные степени больших чисел? Что значит быстрый алгоритм вычисления и что такое сложная вычислительная задача? Где взять большие простые числа? Как, например, построить простое число в 65 десятичных знаков? Существуют ли другие способы решения сравнения (2)? Ведь, если можно найти решение (2), не вычисляя секретный показатель d или не разлагая число m на простые сомножители, да ещё сделать это достаточно быстро, вся система RSA разваливается. Наверное, читателю могут прийти в голову и другие вопросы.

Начнем с конца. За 17 лет, прошедших между публикациями работ [2] и [6], никто так и не смог расшифровать предложенную авторами RSA фразу. Конечно, это всего лишь косвенное подтверждение стойкости системы RSA, но все же достаточно убедительное. Ниже мы обсудим теоретические проблемы, возникающие при решении полиномиальных сравнений.

Мы не будем обсуждать, как выполнять арифметические действия с большими целыми числами, рекомендуем читателю обратиться к замечательной книжке Д. Кнута [7, гл. 4]. Заметим только, что большое число всегда можно разбить на меньшие блоки, с которыми компьютер может оперировать так же, как мы оперируем с цифрами, когда проводим вычисления вручную на бумаге. Конечно, для этого нужны специальные программы. Созданы и получили достаточно широкое распространение даже специальные языки программирования для вычислений с большими числами. Укажем здесь два из них — PARI и UBASIC. Эти языки свободно распространяются. Информацию о том, как их получить в пользование, можно найти в книге [19].

2. Сложность теоретико-числовых алгоритмов

Сложность алгоритмов теории чисел обычно принято измерять количеством арифметических операций (сложений, вычитаний, умножений и делений с остатком), необходимых для выполнения всех действий, предписанных алгоритмом. Впрочем, это определение не учитывает величины чисел, участвующих в вычислениях. Ясно, что перемножить два стозначных числа значительно сложнее, чем два однозначных, хотя при этом и в том, и в другом случае выполняется лишь одна арифметическая операция. Поэтому иногда учитывают ещё и величину чисел, сводя дело к так называемым битовым операциям, т. е. оценивая количество необходимых операций с цифрами 0 и 1, в двоичной записи чисел. Это зависит от рассматриваемой задачи, от целей автора и т.д.

На первый взгляд странным также кажется, что операции умножения и деления приравниваются по сложности к операциям сложения и вычитания. Житейский опыт подсказывает, что умножать числа значительно сложнее, чем складывать их. В действительности же, вычисления можно организовать так, что на умножение или деление больших чисел понадобится не намного больше битовых операций, чем на сложение. В книге [8] описывается алгоритм Шёнхаге – Штрассена, основанный на так называемом быстром преобразовании Фурье, и требующий $O(n \ln n \operatorname{lnln} n)$ битовых операций для умножения двух n-разрядных двоичных чисел. Таким же количеством битовых операций можно обойтись при выполнении деления с остатком двух двоичных чисел, записываемых не более, чем n цифрами. Для сравнения отметим, что сложение n-разрядных двоичных чисел требует $O(n)$ битовых операций.

Говоря в этой статье о сложности алгоритмов, мы будем иметь в виду количество арифметических операций. При построении эффективных алгоритмов и обсуждении верхних оценок сложности обычно хватает интуитивных понятий той области математики, которой принадлежит алгоритм. Формализация же этих понятий требуется лишь тогда, когда речь идёт об отсутствии алгоритма или доказательстве нижних оценок сложности. Более детальное и формальное обсуждение этих вопросов см. в статье [9].

Приведем теперь примеры достаточно быстрых алгоритмов с оценками их сложности. Здесь и в дальнейшем мы не будем придерживаться формального описания алгоритмов, стараясь в первую очередь объяснить смысл выполняемых действий.

Следующий алгоритм вычисляет $a^d \pmod m$ за $O(\ln m)$ арифметических операций. При этом, конечно, предполагается, что натуральные числа a и d не превосходят по величине m.

1. Алгоритм вычисления $a^d \pmod m$.

1) Представим d в двоичной системе счисления $d = d_0 2^r + \cdots + d_{r-1}2 + d_r$, где d_i, цифры в двоичном представлении, равны 0 или 1, $d_0 = 1$.

2) Положим $a_0 = a$ и затем для $i = 1, \ldots, r$ вычислим

$$a_i \equiv a_{i-1}^2 \cdot a^{d_i} \pmod m.$$

3) a_r есть искомый вычет $a^d \pmod m$.

Справедливость этого алгоритма вытекает из сравнения

$$a_i \equiv a^{d_0 2^i + \cdots + d_i} \pmod m,$$

легко доказываемого индукцией по i.

Так как каждое вычисление на шаге 2 требует не более трёх умножений по модулю m и этот шаг выполняется $r \leqslant \log_2 m$ раз, то сложность алгоритма может быть оценена величиной $O(\ln m)$.

Второй алгоритм — это классический алгоритм Евклида вычисления наибольшего общего делителя целых чисел. Мы предполагаем заданными два натуральных числа a и b и вычисляем их наибольший общий делитель (a, b).

2. Алгоритм Евклида.

1) Вычислим r — остаток от деления числа a на b, $a = bq + r$, $0 \leqslant r < b$.

2) Если $r = 0$, то b есть искомое число.

3) Если $r \neq 0$, то заменим пару чисел $\langle a, b\rangle$ парой $\langle b, r\rangle$ и перейдем к шагу 1.

Не останавливаясь на объяснении, почему алгоритм действительно находит (a, b), это общеизвестно, докажем некоторую оценку его сложности.

Теорема 1. *При вычислении наибольшего общего делителя (a, b) с помощью алгоритма Евклида будет выполнено не более $5p$ операций деления с остатком, где p есть количество цифр в десятичной записи меньшего из чисел a и b.*

Доказательство. Положим $r_0 = a > b$ и определим $r_1, r_2, \ldots, r_n$ — последовательность делителей, появляющихся в процессе выполнения шага 1 алгоритма Евклида. Тогда

$$r_1 = b, \ldots, \quad 0 \leqslant r_{i+1} < r_i, \quad i = 0, 1, \ldots n-1.$$

Пусть также $u_0 = 1$, $u_1 = 1$, $u_{k+1} = u_k + u_{k-1}$, $k \geqslant 1$, — последовательность Фибоначчи. Индукцией по i от $i = n-1$ до $i = 0$ легко доказывается неравенство $r_{i+1} \geqslant u_{n-i}$. А так как $u_n \geqslant 10^{(n-1)/5}$, то имеем неравенства $10^p > b = r_1 \geqslant u_n \geqslant 10^{(n-1)/5}$ и $n < 5p + 1$.

Немного подправив алгоритм Евклида, можно достаточно быстро решать сравнения $ax \equiv 1 \pmod{b}$ при условии, что $(a, b) = 1$. Эта задача равносильна поиску целых решений уравнения $ax + by = 1$.

3. Алгоритм решения уравнения $ax + by = 1$.

0) Определим матрицу $E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

1) Вычислим r *— остаток от деления числа* a *на* b, $a = bq + r$, $0 \leqslant r < b$.

2) Если $r = 0$, *то второй столбец матрицы* E *даёт вектор* $\begin{pmatrix} x \\ y \end{pmatrix}$ *решений уравнения.*

3) Если $r \neq 0$, *то заменим матрицу* E *матрицей* $E \cdot \begin{pmatrix} 0 & 1 \\ 1 & -q \end{pmatrix}$.

4) Заменим пару чисел $\langle a, b\rangle$ *парой* $\langle b, r\rangle$ *и перейдем к шагу 1.*

Если обозначить через E_k матрицу E, возникающую в процессе работы алгоритма перед шагом 2 после k делений с остатком (шаг 1), то в обозначениях из доказательства теоремы 1 в этот момент выполняется векторное равенство $\langle a, b\rangle \cdot E_k = \langle r_{k-1}, r_k\rangle$. Его легко доказать индукцией по k. Поскольку числа a и b взаимно просты, имеем $r_n = 1$, и это доказывает, что алгоритм действительно даёт решение уравнения $ax + by = 1$. Буквой n мы обозначили количество делений с остатком, которое в точности такое же, как и в алгоритме Евклида.

Три приведённых выше алгоритма относятся к разряду так называемых полиномиальных алгоритмов. Это название носят алгоритмы, сложность которых оценивается сверху степенным образом в зависимости от длины записи входящих чисел (см. подробности в [9]). Если наибольшее из чисел, подаваемых на вход алгоритма, не превосходит m, то сложность алгоритмов этого типа оценивается величиной $O(\ln^c m)$, где c — некоторая абсолютная постоянная. Во всех приведённых выше примерах $c = 1$.

Полиномиальные алгоритмы в теории чисел — большая редкость. Да и оценки сложности алгоритмов чаще всего опираются на какие-либо не доказанные, но правдоподобные гипотезы, обычно относящиеся к аналитической теории чисел.

Для некоторых задач эффективные алгоритмы вообще не известны. Иногда в таких случаях все же можно предложить последовательность действий, которая, «если повезет», быстро приводит к требуемому результату. Существует класс так называемых вероятностных алгоритмов, которые дают правильный результат, но имеют вероятностную оценку времени работы. Обычно работа этих алгоритмов зависит от одного или нескольких параметров. В худшем случае они работают достаточно

долго. Но удачный выбор параметра определяет быстрое завершение работы. Такие алгоритмы, если множество «хороших» значений параметров велико, на практике работают достаточно эффективно, хотя и не имеют хороших оценок сложности.

Мы будем иногда использовать слова детерминированный алгоритм, чтобы отличать алгоритмы в обычном смысле от вероятностных алгоритмов.

Как пример, рассмотрим вероятностный алгоритм, позволяющий эффективно находить решения полиномиальных сравнений по простому модулю. Пусть p — простое число, которое предполагается большим, и $f(x) \in \mathbb{Z}[x]$ — многочлен, степень которого предполагается ограниченной. Задача состоит в отыскании решений сравнения

$$f(x) \equiv 0 \pmod{p}. \tag{8}$$

Например, речь может идти о решении квадратичных сравнений, если степень многочлена $f(x)$ равна 2. Другими словами, мы должны отыскать в поле $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ все элементы, удовлетворяющие уравнению $f(x) = 0$.

Согласно малой теореме Ферма, все элементы поля $\mathbb{F}_p$ являются однократными корнями многочлена $x^p - x$. Поэтому, вычислив наибольший общий делитель $d(x) = (x^p - x, f(x))$, мы найдем многочлен $d(x)$, множество корней которого в поле $\mathbb{F}_p$ совпадает с множеством корней многочлена $f(x)$, причем все эти корни однократны. Если окажется, что многочлен $d(x)$ имеет нулевую степень, т. е. лежит в поле $\mathbb{F}_p$, это будет означать, что сравнение (8) не имеет решений.

Для вычисления многочлена $d(x)$ удобно сначала вычислить многочлен $c(x) \equiv x^p \pmod{f(x)}$, пользуясь алгоритмом, подобным описанному выше алгоритму возведения в степень (напомним, что число p предполагается большим). А затем с помощью аналога алгоритма Евклида вычислить $d(x) = (c(x) - x, f(x))$. Всё это выполняется за полиномиальное количество арифметических операций.

Таким образом, обсуждая далее задачу нахождения решений сравнения (8), мы можем предполагать, что в кольце многочленов $\mathbb{F}_p[x]$ справедливо равенство

$$f(x) = (x - a_1) \cdot \ldots \cdot (x - a_n), \quad a_i \in \mathbb{F}_p, a_i \neq a_j.$$

4. Алгоритм нахождения делителей многочлена $f(x)$ в кольце $\mathbb{F}_p[x]$.

1) Выберем каким-либо способом элемент $\delta \in \mathbb{F}_p$.

2) Вычислим наибольший общий делитель $g(x) = (f(x), (x+\delta)^{\frac{p-1}{2}} - 1)$.

3) Если многочлен $g(x)$ окажется собственным делителем $f(x)$, то

многочлен $f(x)$ распадётся на два множителя и с каждым из них независимо нужно будет проделать все операции, предписываемые настоящим алгоритмом для многочлена $f(x)$.

4) Если окажется, что $g(x) = 1$ или $g(x) = f(x)$, следует перейти к шагу 1 и, выбрав новое значение δ, продолжить выполнение алгоритма.

Количество операций на шаге 2 оценивается величиной $O(\ln p)$, если вычисления проводить так, как это указывалось выше при нахождении $d(x)$. Выясним теперь, сколь долго придётся выбирать числа δ, пока на шаге 2 не будет найден собственный делитель $f(x)$.

Количество решений уравнения $(t+a_1)^{\frac{p-1}{2}} = (t+a_2)^{\frac{p-1}{2}}$ в поле $\mathbb{F}_p$ не превосходит $\frac{p-3}{2}$. Это означает, что подмножество $D \subset \mathbb{F}_p$, состоящее из элементов δ, удовлетворяющих условиям

$$(\delta+a_1)^{\frac{p-1}{2}} \neq (\delta+a_2)^{\frac{p-1}{2}}, \quad \delta \neq -a_1,\ \delta \neq -a_2,$$

состоит не менее, чем из $\frac{p-1}{2}$ элементов. Учитывая теперь, что каждый ненулевой элемент $b \in \mathbb{F}_p$ удовлетворяет одному из равенств $b^{\frac{p-1}{2}} = 1$, либо $b^{\frac{p-1}{2}} = -1$, заключаем, что для $\delta \in D$ одно из чисел a_1, a_2 будет корнем многочлена $(x+\delta)^{\frac{p-1}{2}} - 1$, а другое — нет. Для таких элементов δ многочлен $g(x)$, определённый на шаге 2 алгоритма, будет собственным делителем многочлена $f(x)$.

Итак, существует не менее $\frac{p-1}{2}$ «удачных» выборов элемента δ, при которых на шаге 2 алгоритма многочлен $f(x)$ распадётся на два собственных множителя. Следовательно, при «случайном» выборе элемента $\delta \in \mathbb{F}_p$, вероятность того, что многочлен не разложится на множители после k повторений шагов алгоритма 1–4, не превосходит 2^{-k}. Вероятность с ростом k убывает очень быстро. И действительно, на практике этот алгоритм работает достаточно эффективно.

Заметим, что при оценке вероятности мы использовали только два корня многочлена $f(x)$. При $n > 2$ эта вероятность, конечно, ещё меньше. Более тонкий анализ с использованием оценок А. Вейля для сумм характеров показывает, что вероятность для многочлена $f(x)$ не распасться на множители при однократном проходе шагов алгоритма 1–4, не превосходит $2^{-n} + O(p^{-1/2})$. Здесь постоянная в $O(\cdot)$ зависит от n. Детали доказательства см. в [26]. В настоящее время известно элементарное доказательство оценки А. Вейля (см. [11]).

В книге [7] описывается принадлежащий Берлекэмпу детерминированный алгоритм решения сравнения (8), требующий $O(pn^3)$ арифметических операций. Ясно, что он практически бесполезен при больших p, а вот при маленьких p и не очень больших n он работает не очень долго.

Если в сравнении (8) заменить простой модуль p составным модулем m, то задача нахождения решений соответствующего сравнения становится намного более сложной. Известные алгоритмы её решения основаны на сведении сравнения к совокупности сравнений (8) по простым модулям — делителям m, и, следовательно, они требуют разложения числа m на простые сомножители, что, как уже указывалось, является достаточно трудоемкой задачей.

3. Как отличить составное число от простого

Существует довольно эффективный способ убедиться, что заданное число является составным, не разлагая это число на множители. Согласно малой теореме Ферма, если число N простое, то для любого целого a, не делящегося на N, выполняется сравнение

$$a^{N-1} \equiv 1 \pmod{N}. \tag{9}$$

Если же при каком-то a это сравнение нарушается, можно утверждать, что N — составное. Проверка (9) не требует больших вычислений, это следует из алгоритма 1. Вопрос только в том, как найти для составного N целое число a, не удовлетворяющее (9). Можно, например, пытаться найти необходимое число a, испытывая все целые числа подряд, начиная с 2. Или попробовать выбирать эти числа случайным образом на отрезке $1 < a < N$.

К сожалению, такой подход не всегда даёт то, что хотелось бы. Имеются составные числа N, обладающие свойством (9) для любого целого a с условием $(a, N) = 1$. Такие числа называются числами Кармайкла. Рассмотрим, например, число $561 = 3 \cdot 11 \cdot 17$. Так как 560 делится на каждое из чисел 2, 10, 16, то с помощью малой теоремы Ферма легко проверить, что 561 есть число Кармайкла. Можно доказать (Carmichael, 1912), что любое из чисел Кармайкла имеет вид $N = p_1 \cdot \ldots \cdot p_r$, $r \geqslant 3$, где все простые p_i различны, причем $N - 1$ делится на каждую разность $p_i - 1$. Лишь недавно, см. [12], была решена проблема о бесконечности множества таких чисел.

В 1976 г. Миллер предложил заменить проверку (9) проверкой несколько иного условия. Детали последующего изложения можно найти в [10]. Если N — простое число, $N - 1 = 2^s \cdot t$, где t нечётно, то согласно малой теореме Ферма для каждого a с условием $(a, N) = 1$ хотя бы одна из скобок в произведении

$$(a^t - 1)(a^t + 1)(a^{2t} + 1) \cdot \ldots \cdot (a^{2^{s-1}t} + 1) = a^{N-1} - 1$$

делится на N. Обращение этого свойства можно использовать, чтобы отличать составные числа от простых.

Пусть N — нечётное составное число, $N-1=2^s\cdot t$, где t нечётно. Назовем целое число a, $1<a<N$, «хорошим» для N, если нарушается одно из двух условий:
α) N не делится на a;
β) $a^t\equiv 1 \pmod N$ или существует целое k, $0\leqslant k<s$, такое, что

$$a^{2^k t}\equiv -1 \pmod N.$$

Из сказанного ранее следует, что для простого числа N не существует хороших чисел a. Если же N составное число, то, как доказал Рабин, их существует не менее $\frac{3}{4}(N-1)$.

Теперь можно построить вероятностный алгоритм, отличающий составные числа от простых.

5. Алгоритм, доказывающий непростоту числа.

1) Выберем случайным образом число a, $1<a<N$, и проверим для этого числа указанные выше свойства α) и β).

2) Если хотя бы одно из них нарушается, то число N составное.

3) Если выполнены оба условия α) и β), возвращаемся к шагу 1.

Из сказанного выше следует, что составное число не будет определено как составное после однократного выполнения шагов 1–3 с вероятностью не большей 4^{-1}. А вероятность не определить его после k повторений не превосходит 4^{-k}, т. е. убывает очень быстро.

Миллер предложил детерминированный алгоритм определения составных чисел, имеющий сложность $O(\ln^3 N)$, однако справедливость его результата зависит от недоказанной в настоящее время так называемой расширенной гипотезы Римана. Согласно этому алгоритму достаточно проверить условия α) и β) для всех целых чисел a, $2\leqslant a\leqslant 70\ln^2 N$. Если при каком-нибудь a из указанного промежутка нарушается одно из условий α) или β), число N составное. В противном случае оно будет простым или степенью простого числа. Последняя возможность, конечно, легко проверяется.

Напомним некоторые понятия, см. [5], необходимые для формулировки расширенной гипотезы Римана. Они понадобятся нам и в дальнейшем. Пусть $m\geqslant 2$ — целое число. Функция $\chi:\mathbb{Z}\to\mathbb{C}$ называется характером Дирихле по модулю m, или просто характером, если эта функция периодична с периодом m, отлична от нуля только на числах, взаимно простых с m, и мультипликативна, т. е. для любых целых u,v выполняется равенство $\chi(uv)=\chi(u)\chi(v)$. Для каждого m существует ровно $\varphi(m)$

характеров Дирихле. Они образуют группу по умножению. Единичным элементом этой группы является так называемый главный характер χ_0, равный 1 на всех числах, взаимно простых с m, и 0 на остальных целых числах. Порядком характера называется его порядок как элемента мультипликативной группы характеров.

С каждым характером может быть связана так называемая L-функция Дирихле — функция комплексного переменного s, определённая рядом $L(s,\chi)=\sum\limits_{n=1}^{\infty}\dfrac{\chi(n)}{n^s}$. Сумма этого ряда аналитична в области $\operatorname{Re} s > 1$ и может быть аналитически продолжена на всю комплексную плоскость. Следующее соотношение $L(s,\chi_0)=\zeta(s)\prod\limits_{p|m}(1-p^{-s})$ связывает L-функцию, отвечающую главному характеру, с дзета-функцией Римана $\zeta(s)=\sum\limits_{n=0}^{\infty}\dfrac{1}{n^s}$. Расширенная гипотеза Римана утверждает, что комплексные нули всех L-функций Дирихле, расположенные в полосе $0<\operatorname{Re} s<1$, лежат на прямой $\operatorname{Re} s=\frac{1}{2}$. В настоящее время не доказана даже простейшая форма этой гипотезы — классическая гипотеза Римана, утверждающая такой же факт о нулях дзета-функции.

В 1952 г. Анкени с помощью расширенной гипотезы Римана доказал, что для каждого простого числа q существует квадратичный невычет a, удовлетворяющий неравенствам $2\leqslant a\leqslant 70\ln^2 q$. Константа 70 была сосчитана позднее. Именно это утверждение и лежит в основе алгоритма Миллера. В 1957 г. Берджесс доказал существование такого невычета без использования расширенной гипотезы Римана, но с худшей оценкой $2\leqslant a\leqslant q^{\frac{1}{4\sqrt{e}}}+\varepsilon$, справедливой при любом положительном ε и q, большем некоторой границы, зависящей от ε.

Алгоритм Миллера принципиально отличается от алгоритма 5, так как полученное с его помощью утверждение о том, что число N — составное, опирается на недоказанную расширенную гипотезу Римана и потому может быть неверным. В то время как вероятностный алгоритм 5 даёт совершенно правильный ответ для составных чисел. Несмотря на отсутствие оценок сложности, на практике он работает вполне удовлетворительно.

4. Как строить большие простые числа

Мы не будем описывать здесь историю этой задачи, рекомендуем обратиться к книге [7] и обзорам [10, 11]. Конечно же, большие простые числа можно строить сравнительно быстро. При этом можно обеспечить

их случайное распределение в заданном диапазоне величин. В противном случае теряла бы всякий практический смысл система шифрования RSA. Наиболее эффективным средством построения простых чисел является несколько модифицированная малая теорема Ферма.

Теорема 2. *Пусть N, S — нечётные натуральные числа, $N-1 = S \cdot R$, причем для каждого простого делителя q числа S существует целое число a такое, что*

$$a^{N-1} \equiv 1 \pmod{N}, \quad (a^{\frac{N-1}{q}} - 1, N) = 1. \tag{10}$$

Тогда каждый простой делитель p числа N удовлетворяет сравнению

$$p \equiv 1 \pmod{2S}.$$

Доказательство. Пусть p — простой делитель числа N, а q — некоторый делитель S. Из условий (10) следует, что в поле вычетов $\mathbb{F}_p$ справедливы соотношения

$$a^{N-1} = 1, \ a^{\frac{N-1}{q}} \neq 1, \ a^{p-1} = 1. \tag{11}$$

Обозначим буквой r порядок элемента a в мультипликативной группе поля $\mathbb{F}_p$. Первые два из соотношений (11) означают, что q входит в разложение на простые множители числа r в степени такой же, как и в разложение $N-1$, а последнее — что $p-1$ делится на r. Таким образом, каждый простой делитель числа S входит в разложение $p-1$ в степени не меньшей, чем в S, так что $p-1$ делится на S. Кроме того, $p-1$ чётно. Теорема 2 доказана.

Следствие. *Если выполнены условия теоремы 2 и $R \leqslant 4S+2$, то N — простое число.*

Действительно, пусть N равняется произведению не менее двух простых чисел. Каждое из них, согласно утверждению теоремы 2, не меньше, чем $2S+1$. Но тогда $(2S+1)^2 \leqslant N = SR+1 \leqslant 4S^2+2S+1$. Противоречие и доказывает следствие.

Покажем теперь, как с помощью последнего утверждения, имея большое простое число S, можно построить существенно большее простое число N. Выберем для этого случайным образом чётное число R на промежутке $S \leqslant R \leqslant 4S+2$ и положим $N = SR+1$. Затем проверим число N на отсутствие малых простых делителей, разделив его на малые простые числа; испытаем N некоторое количество раз с помощью алгоритма 5. Если при этом выяснится, что N — составное число, следует выбрать новое значение R и опять повторить вычисления. Так следует делать до

тех пор, пока не будет найдено число N, выдержавшее испытания алгоритмом 5 достаточно много раз. В этом случае появляется надежда на то, что N — простое число, и следует попытаться доказать простоту с помощью тестов теоремы 2.

Для этого можно случайным образом выбирать число a, $1 < a < N$, и проверять для него выполнимость соотношений

$$a^{N-1} \equiv 1 \pmod{N}, \quad (a^R - 1, N) = 1. \tag{12}$$

Если при выбранном a эти соотношения выполняются, то, согласно следствию из теоремы 2, можно утверждать, что число N простое. Если же эти условия нарушаются, нужно выбрать другое значение a и повторять эти операции до тех пор, пока такое число не будет обнаружено.

Предположим, что построенное число N действительно является простым. Зададимся вопросом, сколь долго придётся перебирать числа a, пока не будет найдено такое, для которого будут выполнены условия (12). Заметим, что для простого числа N первое условие (12), согласно малой теореме Ферма, будет выполняться всегда. Те же числа a, для которых нарушается второе условие (12), удовлетворяют сравнению $a^R \equiv 1 \pmod{N}$. Как известно, уравнение $x^R = 1$ в поле вычетов $\mathbb{F}_N$ имеет не более R решений. Одно из них $x = 1$. Поэтому на промежутке $1 < a < N$ имеется не более $R-1$ чисел, для которых не выполняются условия (12). Это означает, что, выбирая случайным образом числа a на промежутке $1 < a < N$, при простом N можно с вероятностью большей, чем $1-O(S^{-1})$, найти число a, для которого будут выполнены условия теоремы 2, и тем доказать, что N действительно является простым числом.

Заметим, что построенное таким способом простое число N будет удовлетворять неравенству $N > S^2$, т. е. будет записываться вдвое большим количеством цифр, чем исходное простое число S. Заменив теперь число S на найденное простое число N и повторив с этим новым S все указанные выше действия, можно построить ещё большее простое число. Начав с какого-нибудь простого числа, скажем, записанного 10 десятичными цифрами (простоту его можно проверить, например, делением на маленькие табличные простые числа), и повторив указанную процедуру достаточное число раз, можно построить простые числа нужной величины.

Обсудим теперь некоторые теоретические вопросы, возникающие в связи с нахождением числа R, удовлетворяющего неравенствам $S \leqslant R \leqslant 4S + 2$, и такого, что $N = SR + 1$ — простое число. Прежде всего, согласно теореме Дирихле, доказанной ещё в 1839 г., прогрессия $2Sn + 1$, $n = 1, 2, 3, \ldots$ содержит бесконечное количество простых чисел. Нас инте-

ресуют простые числа, лежащие недалеко от начала прогрессии. Оценка наименьшего простого числа в арифметической прогрессии была получена в 1944 г. Ю. В. Линником. Соответствующая теорема утверждает, что наименьшее простое число в арифметической прогрессии $2Sn+1$ не превосходит S^C, где C — некоторая достаточно большая абсолютная постоянная. В предположении справедливости расширенной гипотезы Римана можно доказать, [13, стр. 272], что наименьшее такое простое число не превосходит $c(\varepsilon)\cdot S^{2+\varepsilon}$ при любом положительном ε.

Таким образом, в настоящее время никаких теоретических гарантий для существования простого числа $N=SR+1$, $S\leqslant R\leqslant 4S+2$ не существует. Тем не менее опыт вычислений на ЭВМ показывает, что простые числа в арифметической прогрессии встречаются достаточно близко к её началу. Упомянем в этой связи гипотезу о существовании бесконечного количества простых чисел q с условием, что число $2q+1$ также простое, т. е. простым является уже первый член прогрессии.

Очень важен в связи с описываемым методом построения простых чисел также вопрос о расстоянии между соседними простыми числами в арифметической прогрессии. Ведь убедившись, что при некотором R число $N=SR+1$ составное, можно следующее значение R взять равным $R+2$ и действовать так далее, пока не будет найдено простое число N. И если расстояние между соседними простыми числами в прогрессии велико, нет надежды быстро построить нужное число N. Перебор чисел R до того момента, как мы наткнемся на простое число N окажется слишком долгим. В более простом вопросе о расстоянии между соседними простыми числами p_n и p_{n+1} в натуральном ряде доказано лишь, что $p_{n+1}-p_n=O\left(p_n^{\frac{38}{61}+\varepsilon}\right)$, что, конечно, не очень хорошо для наших целей. Вместе с тем существует так называемая гипотеза Крамера (1936 г.), что $p_{n+1}-p_n=O(\ln^2 p_n)$, дающая вполне приемлемую оценку. Примерно такой же результат следует и из расширенной гипотезы Римана. Вычисления на ЭВМ показывают, что простые числа в арифметических прогрессиях расположены достаточно плотно.

В качестве итога обсуждения в этом пункте подчеркнём следующее: если принять на веру, что наименьшее простое число, а также расстояние между соседними простыми числами в прогрессии $2Sn+1$ при $S\leqslant n\leqslant 4S+2$ оцениваются величиной $O(\ln^2 S)$, то описанная схема построения больших простых чисел имеет полиномиальную оценку сложности. Кроме того, несмотря на отсутствие теоретических оценок времени работы алгоритмов, отыскивающих простые числа в арифметических прогрессиях со сравнительно большой разностью, на практике эти алгоритмы

работают вполне удовлетворительно. На обычном персональном компьютере без особых затрат времени строятся таким способом простые числа порядка 10^{300}.

Конечно, способ конструирования простых чисел для использования в схеме RSA должен быть массовым, а сами простые числа должны быть в каком-то смысле хорошо распределёнными. Это вносит ряд дополнительных осложнений в работу алгоритмов. Впрочем, описанная схема допускает массу вариаций. Все эти вопросы рассматриваются в статье [14].

Наконец, отметим, что существуют методы построения больших простых чисел, использующие не только простые делители $N-1$, но и делители чисел $N+1$, N^2+1, $N^2 \pm N+1$. В основе их лежит использование последовательностей целых чисел, удовлетворяющих линейным рекуррентным уравнениям различных порядков. Отметим, что последовательность a^n, члены которой присутствуют в формулировке малой теоремы Ферма, составляет решение рекуррентного уравнения первого порядка $u_{n+1} = au_n$, $u_0 = 1$.

5. Как проверить большое число на простоту

Есть некоторое отличие в постановках задач предыдущего и настоящего пунктов. Когда мы строим простое число N, мы обладаем некоторой дополнительной информацией о нем, возникающей в процессе построения. Например, такой информацией является знание простых делителей числа $N-1$. Эта информация иногда облегчает доказательство простоты N.

В этом пункте мы предполагаем лишь, что нам задано некоторое число N, например, выбранное случайным образом на каком-то промежутке, и требуется установить его простоту, или доказать, что оно является составным. Эту задачу за полиномиальное количество операций решает указанный в п. 3 алгоритм Миллера. Однако, справедливость полученного с его помощью утверждения зависит от недоказанной расширенной гипотезы Римана. Если число N выдержало испытания алгоритмом 5 для 100 различных значений параметра a, то, по-видимому, можно утверждать, что оно является простым с вероятностью большей, чем $1 - 4^{-100}$. Эта вероятность очень близка к единице, однако всё же оставляет некоторую тень сомнения на простоте числа N. В дальнейшем в этом пункте мы будем считать, что заданное число N является простым, а нам требуется лишь доказать это.

В настоящее время известны детерминированные алгоритмы различной сложности для доказательства простоты чисел. Мы остановимся подробнее на одном из них, предложенном в 1983 г. в совместной работе

Адлемана, Померанца и Рамели [15]. Для доказательства простоты или непростоты числа N этот алгоритм требует $(\ln N)^{c \ln\ln\ln N}$ арифметических операций. Здесь c — некоторая положительная абсолютная постоянная. Функция $\ln\ln\ln N$ хоть и медленно, но всё же возрастает с ростом N, поэтому алгоритм не является полиномиальным. Но всё же его практические реализации позволяют достаточно быстро тестировать числа на простоту. Существенные усовершенствования и упрощения в первоначальный вариант алгоритма были внесены в работах Х. Ленстры и А. Коена [16, 17]. Мы будем называть описываемый ниже алгоритм алгоритмом Адлемана – Ленстры.

В основе алгоритма лежит использование сравнений типа малой теоремы Ферма, но в кольцах целых чисел круговых полей, т. е. полей, порождённых над полем $\mathbb{Q}$ числами $\zeta_p = e^{2\pi i/p}$ — корнями из 1. Пусть q — простое нечётное число и c — первообразный корень по модулю q, т. е. образующий элемент мультипликативной группы поля $\mathbb{F}_q$, которая циклична. Для каждого целого числа x, не делящегося на q, можно определить его индекс, $\operatorname{ind}_q x \in \mathbb{Z}/(q-1)\mathbb{Z}$, называемый также *дискретным логарифмом*, с помощью сравнения $x \equiv c^{\operatorname{ind}_q x} \pmod q$. Рассмотрим далее два простых числа p, q с условием, что $q-1$ делится на p, но не делится на p^2.

Следующая функция, определённая на множестве целых чисел,

$$\chi(x) = \begin{cases} 0, & \text{если} \quad q|x, \\ \zeta_p^{\operatorname{ind}_q x}, & \text{если} \quad (x,q) = 1 \end{cases}$$

является характером по модулю q и порядок этого характера равен p. Сумма

$$\tau(\chi) = -\sum_{x=1}^{q-1} \chi(x)\zeta_q^x \in \mathbb{Z}[\zeta_p, \zeta_q]$$

называется суммой Гаусса. Формулируемая ниже теорема 3 представляет собой аналог малой теоремы Ферма, используемый в алгоритме Адлемана – Ленстры.

Теорема 3. *Пусть N — нечётное простое число, $(N, pq) = 1$. Тогда в кольце $\mathbb{Z}[\zeta_p, \zeta_q]$ выполняется сравнение*

$$\tau(\chi)^N \equiv \chi(N)^{-N} \cdot \tau(\chi^N) \pmod{NZ[\zeta_p, \zeta_q]}.$$

Если при каких-либо числах p, q сравнение из теоремы 3 нарушается, можно утверждать, что N составное число. В противном случае, если сравнение выполняется, оно даёт некоторую информацию о возможных простых делителях числа N. Собрав такую информацию для различных

p, q, в конце концов удаётся установить, что N имеет лишь один простой делитель и является простым.

В случае $p = 2$ легко проверить, что сравнение из теоремы 3 равносильно хорошо известному в элементарной теории чисел сравнению

$$q^{\frac{N-1}{2}} \equiv \left(\frac{q}{N}\right) \pmod N, \tag{13}$$

где $\left(\frac{q}{N}\right)$ — так называемый символ Якоби. Хорошо известно также, что последнее сравнение выполняется не только для простых q, но и для любых целых q, взаимно простых с N. Заметим также, что для вычисления символа Якоби существует быстрый алгоритм, основанный на законе взаимности Гаусса и, в некотором смысле, подобный алгоритму Евклида вычисления наибольшего общего делителя. Следующий пример показывает, каким образом выполнимость нескольких сравнений типа (13) даёт некоторую информацию о возможных простых делителях числа N.

Пример (Х. Ленстра). Пусть N — натуральное число, $(N, 6) = 1$, для которого выполнены сравнения

$$a^{\frac{N-1}{2}} \equiv \left(\frac{a}{N}\right) \pmod N, \quad \text{при } a = -1, 2, 3, \tag{14}$$

а кроме того с некоторым целым числом b имеем

$$b^{\frac{N-1}{2}} \equiv -1 \pmod N. \tag{15}$$

Как уже указывалось, при простом N сравнения (14) выполняются для любого a, взаимно простого с N, а сравнение (15) означает, что b есть первообразный корень по модулю N. Количество первообразных корней равно $\varphi(N-1)$, т. е. достаточно велико. Таким образом, число b с условием (15) при простом N может быть найдено достаточно быстро с помощью случайного выбора и последующей проверки (15).

Докажем, что из выполнимости (14–15) следует, что каждый делитель r числа N удовлетворяет одному из сравнений

$$r \equiv 1 \pmod{24} \text{ или } r \equiv N \pmod{24}. \tag{16}$$

Не уменьшая общности, можно считать, что r — простое число. Введем теперь обозначения $N - 1 = u \cdot 2^k$, $r - 1 = v \cdot 2^m$, где u и v — нечётные числа. Из (15) и сравнения $b^{r-1} \equiv 1 \pmod r$ следует, что $m \geqslant k$. Далее, согласно (14), выполняются следующие сравнения

$$\left(\frac{a}{N}\right) = \left(\frac{a}{N}\right)^v \equiv a^{uv2^{k-1}} \pmod r, \qquad \left(\frac{a}{r}\right) = \left(\frac{a}{r}\right)^u \equiv a^{uv2^{m-1}} \pmod r,$$

означающие (в силу того, что символ Якоби может равняться лишь -1 или $+1$), что

$$\left(\frac{a}{N}\right)^{2^{m-k}} = \left(\frac{a}{r}\right).$$

При $m > k$ это равенство означает, что $\left(\frac{a}{r}\right) = 1$ при $a = -1, 2, 3$, и, следовательно, $r \equiv 1 \pmod{24}$. Если же $m = k$, то имеем $\left(\frac{a}{N}\right) = \left(\frac{a}{r}\right)$ и $r \equiv N \pmod{24}$. Этим (16) доказано.

Информация такого рода получается и в случае произвольных простых чисел p и q с указанными выше свойствами.

Опишем (очень грубо) схему алгоритма Адлемана – Ленстры для проверки простоты N:

1) выбираются различные простые числа $p_1, \ldots, p_k$ и различные простые нечётные $q_1, \ldots, q_s$ такие, что

а) для каждого j все простые делители числа $q_j - 1$ содержатся среди $p_1, \ldots, p_k$ и $q_j - 1$ не делятся на квадрат простого числа;

б) $S = 2q_1 \cdot \ldots \cdot q_s > \sqrt{N}$.

2) для каждой пары выбранных чисел p, q проводятся тесты, подобные сравнению из теоремы 3. Если N не удовлетворяет какому-либо из этих тестов — оно составное. В противном случае

3) определяется не очень большое множество чисел, с которыми только и могут быть сравнимы простые делители N. А именно, каждый простой делитель r числа N должен удовлетворять сравнению вида

$$r \equiv N^j \pmod{S}, \quad 0 \leqslant j < T = p_1 \cdot \ldots \cdot p_k.$$

4) проверяется, содержит ли найденное множество делители N. Если при этом делители не обнаружены, утверждается, что N — простое число.

Если число N составное, оно обязательно имеет простой делитель r, меньший $\sqrt{N} < S$, который сам содержится среди возможных остатков. Именно на этом свойстве основано применение пункта 4) алгоритма.

Пример. Если выбрать следующие множества простых чисел

$$\{p\} = \{2, 3, 5, 7\} \text{ и } \{q\} = \{3, 7, 11, 31, 43, 71, 211\},$$

то таким способом удается проверять простоту чисел $N < 8{,}5 \cdot 10^{19}$.

Отметим, что в работе [15] для тестирования использовались не сравнения теоремы 3, а закон взаимности для степенных вычетов и так

называемые суммы Якоби. Сумма Якоби

$$J(\chi_1, \chi_2) = -\sum_{x=2}^{q-1} \chi_1(x)\chi_2(1-x)$$

определяется для двух характеров χ_1, χ_2 по модулю q. Если характеры имеют порядок p, то соответствующая сумма Якоби принадлежит кольцу $\mathbb{Z}[\zeta_p]$. Поскольку числа p, участвующие в алгоритме, сравнительно невелики, то вычисления с суммами Якоби производятся в полях существенно меньшей степени, чем вычисления с суммами Гаусса. Это главная причина, по которой суммы Якоби предпочтительнее для вычислений. При $\chi_1\chi_2 \neq \chi_0$ выполняется классическое соотношение

$$J(\chi_1, \chi_2) = \frac{\tau(\chi_1) \cdot \tau(\chi_2)}{\tau(\chi_1 \cdot \chi_2)},$$

связывающее суммы Гаусса с суммами Якоби и позволяющее переписать сравнение теоремы 3 в терминах сумм Якоби (см. [16]). Так, при $p = 3$ и $q = 7$ соответствующее сравнение, справедливое для простых N, отличных от 2, 3, 7, принимает вид

$$(-3\zeta - 2)^{\left[\frac{N}{3}\right]} \cdot (3\zeta + 1)^{\left[\frac{2N}{3}\right]} \equiv \xi \pmod{N\mathbb{Z}[\zeta]},$$

где $\zeta = e^{2\pi i/3}$ и ξ — некоторый корень кубический из 1.

В 1984 г. в работе [17] было внесено существенное усовершенствование в алгоритм, позволившее освободиться от требования неделимости чисел $q-1$ на квадраты простых чисел. В результате, например, выбрав число $T = 2^4 \cdot 3^2 \cdot 5 \cdot 7 = 5040$ и взяв S равным произведению простых чисел q с условием, что T делится на $q-1$, получим $S > 1{,}5 \cdot 10^{52}$, что позволяет доказывать простоту чисел N, записываемых сотней десятичных знаков. При этом вычисления будут проводиться в полях, порождённых корнями из 1 степеней 16, 9, 5 и 7.

Мой персональный компьютер с процессором Pentium-150, пользуясь реализацией этого алгоритма на языке UBASIC, доказал простоту записываемого 65 десятичными знаками, большего из простых чисел в примере Ривеста, Шамира и Адлемана (см. пункт 1) за 8 секунд. Сравнение этих 8 секунд и 17 лет, потребовавшихся для разложения на множители предложенного в примере числа, конечно, впечатляет.

Отметим, что оценка сложности этого алгоритма представляет собой трудную задачу аналитической теории чисел. Как уже указывалось, количество операций оценивается величиной $(\ln N)^{c \ln\ln\ln N}$. Однако соответствующие числа S и T, возникающие в процессе доказательства, не

могут быть явно указаны в зависимости от N. Доказано лишь существование чисел S и T, для которых достигается оценка. Впрочем, есть вероятностный вариант алгоритма, доказывающий простоту простого числа N с вероятностью большей $1-2^{-k}$ за $O(k(\ln N)^{c\ln\ln\ln N})$ арифметических операций. А в предположении расширенной гипотезы Римана эта оценка сложности может быть получена при эффективно указанных S и T.

6. Как раскладывают составные числа на множители

Мы лишь кратко коснемся этой темы, отсылая читателей к книгам [7, 18, 19]. Среди многих алгоритмов разложения мы выберем ту линию развития, которая привела к разложению числа, предложенного RSA.

Поиском эффективных способов разложения целых чисел на множители занимаются уже очень давно. Эта задача интересовала выдающихся учёных в области теории чисел. Вероятно Ферма был первый, кто предложил представить разлагаемое число N в виде разности квадратов $N = x^2 - y^2$, а затем, вычисляя $(N, x-y)$, попытаться найти нетривиальный делитель N. Он же предложил и способ, позволяющий найти требуемое представление. Если разлагаемое число имеет два не очень отличающиеся по величине множителя, этот способ позволяет определить их быстрее, чем простой перебор делителей. Лежандр обратил внимание на то, что при таком подходе достаточно получить сравнение

$$x^2 \equiv y^2 \pmod N. \tag{17}$$

Конечно, не каждая пара чисел, удовлетворяющих ему, позволяет разложить N на множители. Эйлер и Гаусс предложили некоторые способы нахождения чисел, связанных соотношением (17). Лежандр использовал для этой цели непрерывные дроби.

Напомним, что каждому иррациональному числу ξ может быть поставлена в соответствие бесконечная последовательность целых чисел $[b_0;\, b_1,\, b_2, \ldots]$, называемая его непрерывной дробью. Это сопоставление строится следующим образом

$$x_0 = \xi,\ b_i = [x_i],\ x_{i+1} = \frac{1}{x_i - b_i},\ i = 0,\, 1,\, 2, \ldots.$$

Лежандр доказал, что непрерывная дробь квадратичной иррациональности периодична. Если раскладывать в непрерывную дробь число $\xi = \sqrt{N}$, то возникающие в процессе разложения числа x_i имеют вид $x_i = \dfrac{\sqrt{N} + P_i}{Q_i}$ с целыми P_i, Q_i, причем всегда $0 \leqslant P_i < \sqrt{N}$, $0 < Q_i < 2\sqrt{N}$. С каждой непрерывной дробью можно связать последовательность рациональных

чисел, так называемых подходящих дробей, $\dfrac{A_i}{B_i}$, $i \geqslant 0$, вычисляемых по правилам

$$A_{i+1} = b_{i+1}A_i + A_{i-1},\ B_{i+1} = b_{i+1}B_i + B_{i-1},\ i \geqslant 0,$$
$$A_0 = b_0,\ B_0 = A_{-1} = 1,\ B_{-1} = 0$$

и стремящихся к разлагаемому числу. Если в непрерывную дробь разлагается число $\xi = \sqrt{N}$, то справедливо соотношение

$$A_{i-1}^2 - NB_{i-1}^2 = (-1)^i Q_i, \tag{18}$$

из которого следует

$$A_{i-1}^2 \equiv (-1)^i Q_i \pmod{N}. \tag{19}$$

Заметим, что длина периода разложения в непрерывную дробь числа $\xi = \sqrt{N}$ может быть большой и достигать величин порядка $\sqrt{N}$.

В 1971 г. Шенкс предложил использовать сравнения (19) для конструирования чисел, удовлетворяющих (17). Если вычисления проводить до тех пор, пока при чётном i не получится $Q_i = R^2$ при некотором целом R, то пара чисел $\langle A_{i-1}, R\rangle$ будет удовлетворять (17) и с её помощью можно надеяться получить разложение N на простые множители.

В 1975 г. Моррисон и Бриллхарт стали перемножать сравнения (19) при различных i с тем, чтобы таким способом получить квадрат целого числа в правой части. Этот метод, метод непрерывных дробей, позволил впервые разложить на множители седьмое число Ферма $F_7 = 2^{128} + 1$. Для реализации алгоритма выбирается так называемая база множителей $\{p_1, p_2, \ldots, p_s\}$. В неё входят ограниченные по величине некоторым параметром простые числа такие, что $\left(\dfrac{N}{p_i}\right) = 1$. Последнее условие связано с тем, что, согласно (18), в разложение на простые множители чисел Q_i могут входить лишь те простые, для которых N является квадратичным вычетом.

На первом этапе алгоритма каждое очередное число Q_i делится на все числа $p_1, p_2, \ldots, p_s$ и, если оно не разлагается полностью в произведение степеней этих простых, то отбрасывается. Иначе получается разложение

$$(-1)^i Q_i = (-1)^{a_0} \prod_{j=1}^{s} p_j^{a_j}. \tag{20}$$

Этому номеру i сопоставляется вектор $(a_0, a_1, \ldots, a_s)$ (вектор показателей). Затем вычисляется следующее значение Q_{i+1}, и с ним проделывается в точности такая же процедура.

Эти вычисления проводятся до тех пор, пока не будет построено $s+2$ вектора показателей. В получившейся матрице показателей, очевидно, можно подобрать вектора-строки так, что их сумма будет вектором с чётными координатами $2(b_0, b_1, \ldots b_s)$. Если Δ — множество номеров векторов, вошедших в эту сумму, то, как легко проверить с помощью (19), имеет место сравнение

$$\left(\prod_{i\in\Delta} A_{i-1}\right)^2 \equiv \left(\prod_{j=1}^{s} p_j^{b_j}\right)^2 \pmod N.$$

Если с помощью этого сравнения не удаётся разложить N на множители, разложение в непрерывную дробь продолжается, продолжается набор векторов показателей и т. д.

В этот алгоритм был внесен ряд усовершенствований: вместо $\sqrt{N}$ можно раскладывать в непрерывную дробь число $\sqrt{kN}$, где маленький множитель k подбирается так, чтобы в базу множителей вошли все малые простые; была предложена так называемая стратегия раннего обрыва и т. д. Сложность этого алгоритма была оценена в 1982 г. величиной $O(\exp(\sqrt{1{,}5\cdot \ln N\cdot \operatorname{lnln} N}))$. При выводе этой оценки использовался ряд правдоподобных, но не доказанных гипотез о распределении простых чисел. Получившаяся в оценке функция растет медленнее любой степенной функции. Алгоритмы, сложность которых оценивается подобным образом, получили название субэкспоненциальных (в зависимости от $\ln N$).

В 1982 г. Померанцом был предложен ещё один субэкспоненциальный алгоритм — алгоритм квадратичного решета. Его сложность оценивается такой же функцией, как и в методе непрерывных дробей, но вместо константы 1,5 получена лучшая — 9/8. Обозначим $m = \left[\sqrt{N}\right]$, $Q(x) = (x+m)^2 - N$ и выберем ту же базу множителей, что и в методе непрерывных дробей. При малых целых значениях x величина $Q(x)$ будет сравнительно невелика. Следующий шаг объясняет название алгоритма — квадратичное решето. Вместо того, чтобы перебирать числа x и раскладывать соответствующие значения $Q(x)$ на множители, алгоритм сразу отсеивает негодные значения x, оставляя лишь те, для которых $Q(x)$ имеет делители среди элементов базы множителей.

Задав некоторую границу B, для каждого простого числа p, входящего в базу множителей, и каждого показателя степени a, с условием $p^a \leqslant B$ находим решения x квадратичного сравнения $Q(x) \equiv 0 \pmod{p^a}$. Множество решений обозначим буквой Λ. Итак, для каждого $x \in \Lambda$ найдётся элемент базы множителей, а может быть и не один, входящий в некоторой степени в разложение на простые сомножители числа $Q(x)$. Те числа

x, при которых значения $Q(x)$ оказываются полностью разложенными, дают нам вектор показателей, как и в алгоритме непрерывных дробей. Если таких векторов окажется достаточно много, с ними можно проделать те же операции, что и в алгоритме непрерывных дробей.

Мы кратко описали здесь лишь основную идею алгоритма. Помимо этого, используется много других дополнительных соображений и различных технических приемов. Например, аналог соотношения (20) имеет вид

$$Q(x) = q_1 q_2 (-1)^{a_0} \prod_{j=1}^{s} p_j^{a_j} \pmod N. \tag{21}$$

В нем допускается наличие двух дополнительных больших простых множителей $B_1 < q_i < B_2$. Эти множители впоследствии при перемножении значений $Q(x)$ исключаются.

Некоторые детали реализации алгоритма можно найти в работе [6]. Отметим здесь только, что на множители раскладывалось число $5N$, база множителей состояла из -1 и 524338 простых чисел, меньших, чем $B_1 = 16333609$. При этом было использовано $B_2 = 2^{30}$. В результате просеивания получилось 112011 соотношений вида (21) без множителей q_i, 1431337 соотношений с одним таким множителем и 6881138 соотношений с двумя множителями. Именно на поиск всех этих соотношений понадобились 220 дней и большое количество работавших параллельно компьютеров. На втором шаге алгоритма, когда из соотношений (21) комбинировались чётные векторы показателей степеней, приходилось работать с матрицами, размеры которых измерялись сотнями тысяч битов. Этот второй шаг потребовал 45 часов работы. Уже четвёртый вектор с чётными показателями привёл к искомому разложению на множители.

Заключение

Мы затронули в этой статье лишь небольшую часть вопросов, связанных с теоретико-числовыми алгоритмами и оценками их сложности. За рамками остались даже чрезвычайно важные для криптографии вопросы дискретного логарифмирования, т. е. поиска чисел x, удовлетворяющих сравнению $a \equiv b^x \pmod p$ при заданных целых a, b, p, см. например, [20, 21]. Мы не описывали перспективные исследования, связанные с распространением алгоритмов решета на поля алгебраических чисел (решето числового поля), и использование их для разложения целых чисел на множители или решения задачи дискретного логарифмирования, см. [22, 20, 25]. Именно с помощью этих алгоритмов достигнуты теоретические оценки сложности разложения на множители $\exp(c(\ln N)^{1/3}(\ln\ln N)^{2/3})$. Не

были затронуты эллиптические кривые, т. е. определённые с точностью до обратимого множителя пропорциональности множества точек

$$E_{a,b} = \{(x, y, z) \in (\mathbb{Z}/m\mathbb{Z})^3 | y^2 z = x^3 + axz^2 + bz^3\},$$

обладающие групповой структурой. С их помощью удалось построить весьма эффективные алгоритмы разложения чисел на множители и проверки целых чисел на простоту. В отличие от мультипликативной группы $(\mathbb{Z}/m\mathbb{Z})^*$, порядок группы $E_{a,b}$ при одном и том же m меняется в зависимости от целых параметров a, b. Это оказывается весьма существенным, например, при разложении чисел m на множители. Мы отсылаем читателей за подробностями использования эллиптических кривых к статье [23].

Список литературы

[1] *Ященко В. В.* Основные понятия криптографии // Математическое просвещение. Сер. 3, №2, 1998. С. 53–70.

[2] *Rivest R. L., Shamir A., Adleman L.* A method for obtaining digital signatures and public key cryptosystems // Commun. ACM. V.21, No 2, 1978. P. 120–126.

[3] *Gardner M.* A new kind of cipher that would take millions of years to break // Sci. Amer. 1977. P. 120–124.

[4] *Виноградов И. М.* Основы теории чисел. М.: Наука, 1972.

[5] *Карацуба А. А.* Основы аналитической теории чисел. М.: Наука, 1983 г.

[6] *Atkins D., Graff M., Lenstra A. K. and Leyland P. C.* The magic words are squeamish ossifrage // ASIACRYPT–94, Lect. Notes in Comput. Sci. V. 917. Springer, 1995.

[7] *Кнут Д.* Искусство программирования на ЭВМ. Т.2: Получисленные алгоритмы. М.: Мир, 1977.

[8] *Ахо А., Хопкрофт Дж., Ульман Дж.* Построение и анализ вычислительных алгоритмов. М.: Мир, 1979.

[9] *Варновский Н. П.* Криптография и теория сложности // Математическое просвещение. Сер. 3, №2, 1998. С. 71–86.

[10] *Williams H. C.* Primality testing on a computer // Ars Combin., 5, 1978. P. 127–185. (Русский перевод: Кибернетический сборник, вып. 23, 1986. С. 51–99.)

[11] *Василенко О. Н.* Современные способы проверки простоты чисел // Кибернетический сборник, вып. 25, 1988. С. 162–188.

[12] *Alford W. R., Granville A., Pomerance C.* There are infinitely many Carmichael numbers // Ann. Math. 140, 1994. P. 703–722.

[13] *Прахар К.* Распределение простых чисел. М.: Мир, 1967.

[14] *Plaisted D. A.* Fast verification, testing, and generation of large primes // Theor. Comp. Sci. 9, 1979. P. 1–16.

[15] *Adleman L. M., Pomerance C., Rumely R. S.* On distinguishing prime numbers from composite numbers // Annals of Math. 117, 1983. P. 173–206.

[16] *Lenstra H. W. (jr.)* Primality testing algorithms (after Adleman, Rumely and Williams) // Lecture Notes in Math. V. 901, 1981. P. 243–257.

[17] *Cohen H., Lenstra H. W. (jr.)* Primality testing and Jacobi sums // Math. of Comput. V. 42, №165, 1984. P. 297–330.

[18] *Riesel H.* Prime numbers and computer methods for factorization. Birkhauser, 1985.

[19] *Cohen H.* A course in computational algebraic number theory. Graduate-Texts in Math. V. 138. New York, Springer, 1993.

[20] *Coppersmith D., Odlyzko A. M., Schroeppel R.* Discrete logarithms in $GF(p)$ // Algorithmica. V. 1, 1986. P. 1–15.

[21] *McCurley K. S.* The discrete logarithm problem // Proc. of Symp. in Appl. Math. V. 42, 1990. P. 49–74.

[22] *Lenstra A. K., Lenstra H. W., Manasse M. S., Pollard J. M.* The number field sieve // Proc. 22nd Ann. ACM Symp. on Theory of Computing. Baltimore, May 14–16, 1990. P. 564–572.

[23] *Lenstra H. W. (jr.)* Elliptic curves and number-theoretic algorithms // ICM86. P. 99–120. (Русский перевод: Международный конгресс математиков в Беркли, М.: Мир, 1991, С. 164–193.)

[24] *Koblitz N.* A Course in Number Theory and Cryptography. 2nd ed. Springer, 1994.

[25] *Lenstra A. K., Lenstra H. W. (jr.)* The Development of the Number Field Sieve. Lect. Notes in Math. V. 1554. Springer, 1993.

[26] *Ben-Or M.* Probabilistic algorithms in finite fields. Proc. 22 IEEE Symp. Found. Comp. Sci, 1981. P. 394–398.

Математика разделения секрета

Г. А. Кабатянский*

1. Введение

Рассмотрим следующую, в наше время вполне реальную ситуацию. Два совладельца драгоценности хотят положить её на хранение в сейф. Сейф современный, с цифровым замком на 16 цифр. Так как совладельцы не доверяют друг другу, то они хотят закрыть сейф таким образом, чтобы они могли открыть его вместе, но никак не порознь. Для этого они приглашают третье лицо, называемое дилером, которому они оба доверяют (например, потому что оно не получит больше доступ к сейфу). Дилер случайно выбирает 16 цифр в качестве «ключа», чтобы закрыть сейф, и затем сообщает первому совладельцу втайне от второго первые 8 цифр «ключа», а второму совладельцу втайне от первого — последние 8 цифр «ключа». Такой способ представляется с точки здравого смысла оптимальным, ведь каждый из совладельцев получил «полключа» и что может быть лучше?! Недостатком данного примера является то, что любой из совладельцев, оставшись наедине с сейфом, может за пару минут найти недостающие «полключа» с помощью несложного устройства, перебирающего ключи со скоростью 1 МГц. Кажется, что единственный выход — в увеличении размера «ключа», скажем, вдвое. Но есть другой, математический выход, опровергающий (в данном случае — к счастью) соображения здравого смысла. А именно, дилер независимо выбирает две случайные последовательности по 16 цифр в каждой, сообщает каждому из совладельцев втайне от другого «его» последовательность, а в качестве «ключа», чтобы закрыть сейф, использует последовательность, полученную сложением по модулю 10 соответствующих цифр двух выбранных последовательностей. Довольно очевидно (и ниже мы это докажем), что для каждого из совладельцев все 10^{16} возможных «ключей» одинаково вероятны и остается только перебирать их, что потребует в среднем более полутора лет для устройства, перебирающего ключи со скоростью 100 МГц.

*Работа поддержана Российским фондом фундаментальных исследований (проект №96-01-00884).

И с математической, и с практической точки зрения неинтересно останавливаться на случае двух участников и следует рассмотреть общую ситуацию. Неформально говоря, «схема, разделяющая секрет» (СРС) позволяет «распределить» секрет между n участниками таким образом, чтобы заранее заданные разрешённые множества участников могли однозначно восстановить секрет (совокупность этих множеств называется структурой доступа), а неразрешённые — не получали никакой дополнительной к имеющейся априорной информации о возможном значении секрета. СРС с последним свойством называются совершенными (и только они, как правило, рассматриваются в этой статье).

История СРС начинается с 1979 года, когда эта проблема была поставлена и во многом решена Г. Блейкли [1] и А. Шамиром [2] для случая пороговых (n,k)-СРС (т. е. разрешёнными множествами являются любые множества из k или более элементов). Особый интерес вызвали так называемые идеальные СРС, т. е. такие, где «размер» информации, предоставляемой участнику, не больше «размера» секрета (а меньше, как было показано, он и не может быть). Оказалось [3], что любой такой СРС соответствует матроид (определение, что это такое, см. в п. 4) и, следовательно, не для любой структуры доступа возможно идеальное разделение секрета. С другой стороны, было показано, что для любого набора разрешённых множеств можно построить совершенную СРС, однако известные построения весьма «неэкономны». В данной статье рассматриваются алгебро-геометрические и комбинаторные задачи, возникающие при математическом анализе «схем, разделяющих секрет». Вот пример одной из таких задач.

Будем говорить, что семейство линейных подпространств $\{L_0,\dots,L_n\}$ конечномерного векторного пространства L над полем K удовлетворяет свойству «всё или ничего», если для любого множества $A\subset\{1,\dots,n\}$ линейная оболочка подпространств $\{L_a : a\in A\}$ либо содержит подпространство L_0 целиком, либо пересекается с ним только по вектору $\mathbf{0}$. В п. 3 мы увидим, что такое семейство задаёт «линейную» СРС, у которой множество $A\subset\{1,\dots,n\}$ является разрешённым, если и только если линейная оболочка подпространств $\{L_a : a\in A\}$ содержит подпространство L_0 целиком. В связи с этим понятием возникает ряд вопросов. Например, если поле K конечно ($|K|=q$) и все подпространства $\{L_0,\dots,L_n\}$ одномерны, то каково максимально возможное число участников n для линейных пороговых (n,k)-СРС ($k>1$)? Иначе говоря, каково максимально возможное число векторов $\{h_0,\dots,h_n\}$ таких, что любые k векторов, содержащие вектор h_0, линейно независимы, а любые $k+1$ векторов, содержащие вектор h_0, линейно зависимы. Оказывается, что это свойство

эквивалентно следующему, на первый взгляд более сильному, свойству: любые k векторов линейно независимы, а любые $k+1$ — линейно зависимы. Такие системы векторов изучались в геометрии как N-множества ($N = n+1$) в конечной проективной геометрии $PG(k-1, q)$, в комбинаторике как ортогональные таблицы силы k и индекса $\lambda = 1$, в теории кодирования как проверочные матрицы МДР кодов (подробнее см. [4]). В п. 3 мы приведем известную конструкцию таких множеств с $N = q+1$, а довольно старая гипотеза состоит в том, что это и есть максимально возможное N, за исключением двух случаев: случая $q < k$, когда $N = k+1$, и случая $q = 2^m$, $k = 3$ или $k = q-1$, когда $N = q+2$.

2. Разделение секрета для произвольных структур доступа

Начнем с формальной математической модели. Имеется $n+1$ множество $\mathcal{S}_0, \mathcal{S}_1, \ldots, \mathcal{S}_n$ и (совместное) распределение вероятностей P на их декартовом произведении $\mathcal{S} = \mathcal{S}_0 \times \cdots \times \mathcal{S}_n$. Соответствующие случайные величины обозначаются через S_i. Имеется также некоторое множество Γ подмножеств множества $\{1, \ldots, n\}$, называемое структурой доступа.

Определение 1. Пара $(P, \mathcal{S})$ называется *совершенной вероятностной СРС*, реализующей структуру доступа Γ, если

$$P(S_0 = c_0 | S_i = c_i, i \in A) \in \{0, 1\} \text{ для } A \in \Gamma, \tag{1}$$

$$P(S_0 = c_0 | S_i = c_i, i \in A) = P(S_0 = c_0) \text{ для } A \notin \Gamma. \tag{2}$$

Это определение можно истолковать следующим образом. Имеется множество $\mathcal{S}_0$ всех возможных секретов, из которого секрет s_0 выбирается с вероятностью $p(s_0)$, и имеется СРС, которая «распределяет» секрет s_0 между n участниками, посылая «проекции» $s_1, \ldots, s_n$ секрета с вероятностью $P_{s_0}(s_1, \ldots, s_n)$. Отметим, что i-й участник получает свою «проекцию» $s_i \in \mathcal{S}_i$ и не имеет информации о значениях других «проекций», однако знает все множества $\mathcal{S}_i$, а также оба распределения вероятностей $p(s_0)$ и $P_{s_0}(s_1, \ldots, s_n)$. Эти два распределения могут быть эквивалентно заменены на одно: $P(s_0, s_1, \ldots, s_n) = p(s_0) P_{s_0}(s_1, \ldots, s_n)$, что и было сделано выше. Цель СРС, как указывалось во введении, состоит в том, чтобы:
а) участники из разрешённого множества A (т. е. $A \in \Gamma$) вместе могли бы однозначно восстановить значение секрета — это отражено в свойстве (1);
б) участники, образующие неразрешённое множество A ($A \notin \Gamma$), не могли бы получить дополнительную информацию об s_0, т. е., чтобы вероятность того, что значение секрета $S_0 = c_0$, не зависела от значений «проекций» S_i при $i \in A$ — это свойство (2).

Замечание о терминологии. В англоязычной литературе для обозначения «порции» информации, посылаемой участнику СРС, были введены термины «share» (А. Шамир) и «shadow» (Г. Блейкли). Первый термин оказался наиболее популярным и автор долго боролся с соблазном привлечь массового читателя, постоянно используя в качестве его перевода слово «акция». Неадекватная (во всех смыслах) замена «акции» на «проекцию» может быть несколько оправдана следующим примером.

Пример 1. Множество $\mathcal{S}_0$ всех возможных секретов состоит из 0, 1 и 2, «представленных» соответственно: шаром; кубом, рёбра которого параллельны осям координат; цилиндром, образующие которого параллельны оси Z. При этом диаметры шара и основания цилиндра, и длины ребра куба и образующей цилиндра, равны. Первый участник получает в качестве своей «доли» секрета его проекцию на плоскость XY, а второй — на плоскость XZ. Ясно, что вместе они однозначно восстановят секрет, а порознь — не могут. Однако, эта СРС не является совершенной, так как любой из участников получает информацию о секрете, оставляя только два значения секрета как возможные при данной проекции (например, если проекция — квадрат, то шар невозможен).

Еще одно замечание. Элемент (участник) $x \in \{1, \ldots, n\}$ называется несущественным (относительно Γ), если для любого неразрешённого множества A множество $A \cup x$ также неразрешённое. Очевидно, что несущественные участники настолько несущественны для разделения секрета, что им просто не нужно посылать никакой информации. Поэтому далее, без ограничения общности, рассматриваются только такие структуры доступа Γ, для которых все элементы являются существенными. Кроме того, естественно предполагать, что Γ является монотонной структурой, т. е. из $A \subset B, A \in \Gamma$ следует $B \in \Gamma$.

Пример 2. Рассмотрим простейшую структуру доступа — (n, n)-пороговую схему, т. е. все участники вместе могут восстановить секрет, а любое подмножество участников не может получить дополнительной информации о секрете. Будем строить идеальную СРС, выбирая и секрет, и его проекции из группы Z_q вычетов по модулю q, т. е. $\mathcal{S}_0 = \mathcal{S}_1 = \ldots = \mathcal{S}_n = Z_q$. Дилер генерирует $n-1$ независимых равномерно распределённых на Z_q случайных величин x_i и посылает i-му участнику $(i = 1, \ldots, n-1)$ его «проекцию» $s_i = x_i$, а n-му участнику посылает $s_n = s_0 - (s_1 + \cdots + s_{n-1})$. Кажущееся «неравноправие» n-ого участника тут же исчезает, если мы выпишем распределение $P_{s_0}(s_1, \ldots, s_n)$, которое очевидно равно $1/q^{n-1}$, если $s_0 = s_1 + \cdots + s_n$, и равно 0 — в остальных случаях. Теперь легко проверяется и свойство (2), означающее в дан-

ном случае независимость случайной величины S_0 от случайных величин $\{S_i : i \in A\}$ при любом собственном подмножестве A.

Данное выше определение СРС, оперирующее словами «распределение вероятностей», ниже переведено, почти без потери общности, на комбинаторный язык, который представляется автору более простым для понимания. Произвольная $M \times (n+1)$-матрица V, строки которой имеют вид $\mathbf{v} = (v_0, v_1, \ldots, v_n)$, где $v_i \in \mathcal{S}_i$, называется матрицей комбинаторной СРС, а её строки — «правилами» распределения секрета. Для заданного значения секрета s_0 дилер СРС случайно и равновероятно выбирает строку $\mathbf{v}$ из тех строк матрицы V, для которых значение нулевой координаты равно s_0.

Определение 2. Матрица V задаёт *совершенную комбинаторную СРС*, реализующую структуру доступа Γ, если, во-первых, для любого множества $A \in \Gamma$ нулевая координата любой строки матрицы V однозначно определяется значениями её координат из множества A, и, во-вторых, для любого множества $A \notin \Gamma$ и любых заданных значений координат из множества A число строк матрицы V с данным значением α нулевой координаты не зависит от α.

Сопоставим совершенной вероятностной СРС, задаваемой парой $(P, \mathcal{S})$, матрицу V, состоящую из строк $s \in \mathcal{S}$, таких что $P(s) > 0$. Заметим, что если в определении 1 положить все ненулевые значения P одинаковыми, а условия (1) и (2) переформулировать на комбинаторном языке, то получится определение 2. Это комбинаторное определение несколько обобщается, если допустить в матрице V повторяющиеся строки, что эквивалентно вероятностному определению 1, когда значения вероятностей $P(s)$ — рациональные числа.

Пример 2 (продолжение). Переформулируем данную выше конструкцию (n, n)-пороговой СРС на комбинаторном языке. Строками матрицы V являются все векторы $\mathbf{s}$ такие, что $-s_0 + s_1 + \cdots + s_n = 0$. Очевидно, что V задаёт совершенную комбинаторную СРС для $\Gamma = \{1, \ldots, n\}$, так как для любого собственного подмножества $A \subset \{1, \ldots, n\}$ и любых заданных значений координат из множества A число строк матрицы V с данным значением нулевой координаты равно $q^{n-1-|A|}$.

Удивительно, но простой схемы примера 2 оказывается достаточно, чтобы из неё, как из кирпичиков, построить совершенную СРС для произвольной структуры доступа. А именно, для всех разрешённых множеств, т. е. для $A \in \Gamma$, независимо реализуем описанную только что пороговую $(|A|, |A|)$-СРС, послав тем самым i-му участнику столько «проекций» s_i^A,

скольким разрешённым множествам он принадлежит. Это словесное описание несложно перевести на комбинаторный язык свойств матрицы V и убедиться, что эта СРС совершенна. Как это часто бывает, «совершенная» не значит «экономная», и у данной СРС размер «проекции» оказывается, как правило, во много раз больше, чем размер секрета. Эту схему можно сделать более экономной, так как достаточно реализовать пороговые $(|A|, |A|)$-СРС только для минимальных разрешённых множеств A, т. е. для $A \in \Gamma_{\min}$, где $\Gamma_{\min}$ — совокупность минимальных (относительно включения) множеств из Γ. Тем не менее, для пороговой $(n, n/2)$-СРС размер «проекции» (измеренный, например, в битах) будет в $C_n^{n/2} \sim 2^n/\sqrt{2\pi n}$ раз больше размера секрета (это наихудший случай для рассматриваемой конструкции). С другой стороны, как мы убедимся чуть позже, любая пороговая структура доступа может быть реализована идеально, т. е. при совпадающих размерах «проекции» и секрета. Поэтому естественно возникает вопрос о том, каково максимально возможное превышение размера «проекции» над размером секрета для наихудшей структуры доступа при наилучшей реализации. Формально, $R(n) = \max R(\Gamma)$, где max берётся по всем структурам доступа Γ на n участниках, а $R(\Gamma) = \min \max \frac{\log |S_i|}{\log |S_0|}$, где min берётся по всем СРС, реализующим данную структуру доступа Γ, а max — по $i = 1, \ldots, n$. Приведенная конструкция показывает, что $R(n) \leqslant C_n^{n/2}$. С другой стороны, как было доказано лишь недавно [5], $R(n) \geqslant n/\log n$. Такая огромная «щель» между верхней и нижней оценкой даёт, по нашему мнению, достаточный простор для исследований (автор предполагает, что $R(n)$ растет экспоненциально от n).

3. Линейное разделение секрета.

Начнем с предложенной А. Шамиром [2] элегантной схемы разделения секрета для пороговых структур доступа. Пусть $K = GF(q)$ конечное поле из q элементов (например, $q = p$ — простое число и $K = Z_p$) и $q > n$. Сопоставим участникам n различных ненулевых элементов поля $\{a_1, \ldots, a_n\}$ и положим $a_0 = 0$. При распределении секрета s_0 дилер СРС генерирует $k-1$ независимых равномерно распределённых на $GF(q)$ случайных величин f_j $(j = 1, \ldots, k-1)$ и посылает i-му участнику $(i = 1, \ldots, n)$ «его» значение $s_i = f(a_i)$ многочлена $f(x) = f_0 + f_1 x + \cdots + f_{k-1}x^{k-1}$, где $f_0 = s_0$. Поскольку любой многочлен степени $k-1$ однозначно восстанавливается по его значениям в произвольных k точках (например, по интерполяционной формуле Лагранжа), то любые k участников вместе могут восстановить многочлен $f(x)$ и, следовательно, найти значение секрета как $s_0 = f(0)$. По этой же причине для любых $k-1$ участников,

любых полученных ими значений проекций s_i и любого значения секрета s_0 существует ровно один «соответствующий» им многочлен, т. е. такой, что $s_i = f(a_i)$ и $s_0 = f(0)$. Следовательно, эта схема является совершенной в соответствии с определением 2. «Линейность» данной схемы становится ясна, если записать «разделение секрета» в векторно-матричном виде:

$$\mathbf{s} = \mathbf{f}H, \tag{3}$$

где $\mathbf{s} = (s_0, \ldots, s_n), \mathbf{f} = (f_0, \ldots, f_{k-1})$, $k \times (n+1)$-матрица $H = (h_{ij}) = (a_i^{j-1})$ и $h_{00} = 1$. Заметим, что любые k столбцов этой матрицы линейно независимы, а максимально возможное число столбцов матрицы H равно q, и чтобы добиться обещанного в п. 1 значения $q+1$ надо добавить столбец, соответствующий точке «бесконечность».

Упражнение. *Придумайте сами, как это сделать.*

Возьмём в (3) в качестве H произвольную $r \times (n+1)$-матрицу с элементами из поля K. Получаемую СРС будем называть одномерной линейной СРС. Она является совершенной комбинаторной СРС со структурой доступа Г, состоящей из множеств A таких, что вектор $\mathbf{h}_0$ представим в виде линейной комбинации векторов $\{\mathbf{h}_j : j \in A\}$, где $\mathbf{h}_j$ это j-ый столбец матрицы H. Строками матрицы V, соответствующей данной СРС являются, как видно из (3), линейные комбинации строк матрицы H. Перепишем (3) в следующем виде

$$s_j = (\mathbf{f}, \mathbf{h}_j) \text{ для } j = 0, 1, \ldots, n,$$

где $(\mathbf{f}, \mathbf{h}_j)$ — скалярное произведение векторов $\mathbf{f}$ и $\mathbf{h}_j$. Если $A \in \Gamma$, т. е. если $\mathbf{h}_0 = \sum \lambda_j \mathbf{h}_j$, то

$$s_0 = (\mathbf{f}, \mathbf{h}_0) = (\mathbf{f}, \sum \lambda_j \mathbf{h}_j) = \sum \lambda_j (\mathbf{f}, \mathbf{h}_j) = \sum \lambda_j s_j$$

и, следовательно, значение секрета однозначно находится по его «проекциям». Рассмотрим теперь случай, когда вектор $\mathbf{h}_0$ не представим в виде линейной комбинации векторов $\{\mathbf{h}_j : j \in A\}$. Нам нужно показать, что в этом случае для любых заданных значений координат из множества A число строк матрицы V с данным значением нулевой координаты не зависит от этого значения. В этом нетрудно убедиться, рассмотрев (3) как систему линейных уравнений относительно неизвестных f_i и воспользовавшись тем, что система совместна тогда и только тогда, когда ранг матрицы коэффициентов равен рангу расширенной матрицы, а число решений у совместных систем одинаково и равно числу решений однородной системы.

Указание. Рассмотрите две системы: без «нулевого» уравнения (т. е. со свободным членом) и с ним. Так как вектор $\mathbf{h}_0$ не представим в виде линейной комбинации векторов $\{\mathbf{h}_j : j \in A\}$, то ранг матрицы коэффициентов второй системы на 1 больше ранга матрицы коэффициентов первой системы. Отсюда немедленно следует, что если первая система совместна, то совместна и вторая при любом s_0.

Эта конструкция подводит нас к определению общей линейной СРС. Пусть секрет и его «проекции» представляются как конечномерные векторы $\mathbf{s}_i = (s_i^1, \ldots, s_i^{m_i})$ и генерируются по формуле $\mathbf{s}_i = \mathbf{f}H_i$, где H_i — некоторые $r \times m_i$-матрицы. Сопоставим каждой матрице H_i линейное пространство L_i её столбцов (т. е. состоящее из всех линейных комбинаций вектор-столбцов матрицы H_i). Несложные рассуждения, аналогичные приведённым выше для одномерного случая (все $m_i = 1$), показывают, что данная конструкция даёт совершенную СРС тогда и только тогда, когда семейство линейных подпространств $\{L_0, \ldots, L_n\}$ конечномерного векторного пространства K^r удовлетворяет упомянутому во введении свойству «всё или ничего». При этом множество A является разрешённым ($A \in \Gamma$), если и только если линейная оболочка подпространств $\{L_a : a \in A\}$ содержит подпространство L_0 целиком. С другой стороны, множество A является неразрешённым ($A \notin \Gamma$), если и только если линейная оболочка подпространств $\{L_a : a \in A\}$ пересекается с подпространством L_0 только по вектору $\mathbf{0}$. Отметим, что если бы для некоторого A пересечение L_0 и линейной оболочки $\{L_a : a \in A\}$ было нетривиальным, то участники A не могли бы восстановить секрет однозначно, но получали бы некоторую информацию о нем, т. е. схема не была бы совершенной.

Пример 3. Рассмотрим следующую структуру доступа для случая четырёх участников, задаваемую $\Gamma_{\min} = \{\{1,2\},\{2,3\},\{3,4\}\}$. Она известна как первый построенный пример структуры доступа, для которой не

Таб. 1.

$$H_0 = \begin{bmatrix} 1\,0 \\ 0\,1 \\ 0\,0 \\ 0\,0 \\ 0\,0 \\ 0\,0 \end{bmatrix},\ H_1 = \begin{bmatrix} 0\,0 \\ 0\,0 \\ 1\,0 \\ 0\,1 \\ 0\,0 \\ 0\,0 \end{bmatrix},\ H_2 = \begin{bmatrix} 1\,0\,0 \\ 0\,1\,0 \\ 1\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \\ 0\,0\,0 \end{bmatrix},\ H_3 = \begin{bmatrix} 0\,0\,1 \\ 0\,0\,0 \\ 0\,0\,0 \\ 0\,1\,0 \\ 0\,0\,1 \\ 1\,0\,0 \end{bmatrix},\ H_4 = \begin{bmatrix} 0\,0 \\ 0\,1 \\ 0\,0 \\ 0\,0 \\ 1\,0 \\ 0\,1 \end{bmatrix}.$$

существует идеальной реализации. Более того, было доказано, что для любой её совершенной реализации $R(\Gamma) \geqslant 3/2$. С другой стороны, непосредственная проверка показывает, что выбор матриц $H_0, H_1, \ldots, H_4$, приведенных в таб. 1, даёт совершенную линейную СРС с $R = 3/2$, реализующую эту структуру, которая, следовательно, является и оптимальной (наиболее экономной) СРС.

4. Идеальное разделение секрета и матроиды

Начнем с определения идеальных СРС. Для этого вернемся к комбинаторному определению совершенной СРС. Следующее определение совершенной СРС [3] является даже более общим, чем вероятностное определение 1, поскольку условие (2) заменено в нем на более слабое.

Для произвольного множества $B \subseteq \{0, 1, \ldots, n\}$ обозначим через V_B $M \times |B|$-матрицу, полученную из матрицы V удалением столбцов, номера которых не принадлежат множеству B. Пусть $||W||$ обозначает число различных строк в матрице W.

Определение 3. Матрица V задаёт БД-совершенную СРС, реализующую структуру доступа Γ, если

$$||V_{A\cup 0}|| = ||V_A|| \times ||V_0||^{\delta_\Gamma(A)}, \tag{4}$$

где $\delta_\Gamma(A) = 0$, если $A \in \Gamma$, и $\delta_\Gamma(A) = 1$ в противном случае.

Это определение отличается от определений 1 и 2 тем, что на неразрешённые множества A накладывается довольно слабое условие, а именно, если множество строк V с данными значениями координат из множества A непусто, то все возможные значения секрета встречаются в нулевой координате этих строк (без требований «одинаково часто» как в комбинаторном определении 2 или же «с априорной вероятностью» как в вероятностном определении 1). Легко видеть, что матрица любой совершенной вероятностной СРС задаёт БД-совершенную СРС, но обратное неверно.

Для произвольной комбинаторной СРС, задаваемой матрицей V, определим на множествах $A \subseteq \{0, 1, \ldots, n\}$ функцию $h(A) = \log_q ||V_A||$, где $q = |\mathcal{S}_0|$. Легко проверить, что $\max\{h(A), h(B)\} \leqslant h(A \cup B) \leqslant h(A) + h(B)$ для любых множеств A и B, а условие (4) может быть переписано в виде

$$h_q(V_{A\cup 0}) = h_q(V_A) + \delta_\Gamma(A) h_q(V_0),$$

Лемма. *Для любой БД-совершенной СРС если* $A \notin \Gamma$ *и* $\{A \cup i\} \in \Gamma$, *то* $h(i) \geqslant h(0)$.

Доказательство. По условиям леммы $h(A \cup 0) = h(A) + h(0)$ и $h(A \cup i \cup 0) = h(A \cup i)$. Следовательно,

$$h(A) + h(i) \geqslant h(A \cup i) = h(A \cup i \cup 0) \geqslant h(A \cup 0) = h(A) + h(0). \qquad \blacksquare$$

Так как мы предполагаем, что все точки $i \in \{1, \ldots, n\}$ существенные, т. е. для любого i найдётся подмножество A такое, что $A \notin \Gamma$ и $\{A \cup i\} \in \Gamma$, то из леммы вытекает

Следствие. *Для любой БД-совершенной СРС* $|\mathcal{S}_i| \geqslant |\mathcal{S}_0|$ *для всех* $i = 1, \ldots, n$.

Следствие означает, как мы и предупреждали в начале статьи, что для совершенных СРС «размер» проекции не может быть меньше «размера» секрета. Поэтому БД-совершенная СРС называется идеальной, если $|\mathcal{S}_i| = |\mathcal{S}_0|$ для всех $i = 1, \ldots, n$.

Замечание. Неравенство $|\mathcal{S}_i| \geqslant |\mathcal{S}_0|$ справедливо и для совершенных вероятностных СРС, поскольку их матрицы задают БД-совершенные СРС.

Естественный вопрос состоит в том, для каких структур доступа Γ существуют реализующие их идеальные (вероятностные или комбинаторные) СРС. Как уже отмечалось во введении, наилучший на сегодняшний день ответ использует слово «матроид». Напомним определение матроидов и некоторые их основные свойства (см. [6]).

Матроидом называется конечное множество X и семейство I его подмножеств, называемых независимыми (остальные множества называются зависимыми), если выполнены следующие свойства:

$$\varnothing \in I; \tag{5.1}$$

$$\text{Если } A \in I \text{ и } B \subset A, \text{ то } B \in I; \tag{5.2}$$

$$\text{Если } A, B \in I \text{ и } |A| = |B| + 1,$$
$$\text{то существует } a \in A \backslash B \text{ такое, что } a \cup B \in I. \tag{5.3}$$

Пример 4. Множество X — это множество векторов в некотором линейном векторном пространстве, а независимые подмножества — это линейно независимые подмножества векторов.

Собственно с этого примера и началась теория матроидов, вначале как попытка дать аксиоматическое определение линейной независимости векторов через «внутренние свойства», т. е. не апеллируя к понятию вектора. К счастью, попытка не удалась, так как нашлись матроиды, не представимые как линейные (т. е. как системы векторов), а сама теория матроидов разрослась далеко за пределы «линейной алгебры» (см. [6]).

Пример 5 (матроид Вамоса). Рассмотрим множество $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$ и положим $a = \{1, 2\}$, $b = \{3, 4\}$, $c = \{5, 6\}$ и $d = \{7, 8\}$. Матроид Вамоса определяется как матроид, в котором множества $a \cup c$, $a \cup d$, $b \cup c$, $b \cup d$, $c \cup d$, а также все подмножества из пяти или более элементов являются зависимыми. Известно, что этот матроид не является линейным.

Матроид также можно определить через так называемую ранговую функцию $r(A)$ матроида, определяемую как максимальная мощность независимого подмножества $B \subseteq A$. Очевидно, что независимые множества (и только они) задаются условием $r(A) = |A|$. Ранговая функция матроида обладает свойствами

$$r(A) \in Z,\ r(\varnothing) = 0; \tag{6.1}$$

$$r(A) \leqslant r(A \cup b) \leqslant r(A) + 1; \tag{6.2}$$

$$\text{Если } r(A \cup b) = r(A \cup c) = r(A), \text{ то } r(A \cup b \cup c) = r(A). \tag{6.3}$$

Обратно, пусть некоторая функция $r(A)$ обладает свойствами (6). Назовем независимыми те множества A, для которых $r(A) = |A|$. Тогда эти множества задают матроид, а функция r является его ранговой функцией. Возможно также определить матроид через минимальные зависимые множества, называемые циклами. Матроид называется связным, если для любых двух его точек существует содержащий их цикл.

Теперь мы можем сформулировать основной результат.

Теорема ([3]). *Для любой БД-совершенной идеальной СРС, реализующей структуру доступа* Γ, *независимые множества, определяемые условием* $\log_{|\mathcal{S}_0|} ||V_A|| = |A|$, *задают связный матроид на множестве* $\{0, 1, \ldots, n\}$. *Все циклы этого матроида, содержащие точку* 0, *имеют вид* $0 \cup A$, *где* $A \in \Gamma_{min}$.

Главным в доказательстве теоремы является «проверка» целочисленности функции $h(A)$. В самом деле, $h(\cdot)$ очевидно обладает остальными свойствами (6) и, следовательно, при условии целочисленности является ранговой функцией и задаёт матроид. Доказательство этой теоремы и несколько более общих утверждений можно найти в [7].

Отметим, что из второй части утверждения теоремы следует, что разным идеальным СРС, реализующим данную структуру доступа Γ, всегда соответствует один и тот же матроид, поскольку матроид однозначно определяется всеми циклами, проходящими через фиксированную точку (см. [6]). Тем самым, каждой идеально реализуемой структуре доступа соответствует однозначно определённый матроид.

В связи с теоремой возникает несколько естественных вопросов. Прежде всего, не порождают ли идеальные СРС все матроиды? Нет, например, матроид Вамоса не может быть получен как матроид идеальной СРС [8]. С другой стороны, линейные матроиды есть ни что иное как рассмотренные в п. 3 идеальные одномерные линейные СРС. В связи с этим возникает вопрос о существовании структуры доступа Г, которую невозможно реализовать в виде идеальной одномерной линейной СРС, но можно в виде идеальной многомерной линейной СРС. Недавно такой пример был построен [9], и, значит, мы можем говорить о многомерных линейных матроидах как классе матроидов более общем, чем линейные.

Итак, идеальных СРС больше, чем линейных матроидов, но меньше, чем всех матроидов. Уточнить, «насколько больше», представляется довольно сложной задачей. В частности, существует ли идеально реализуемая структура доступа Г, которую невозможно реализовать как идеальную линейную многомерную СРС?

Список литературы

[1] *Blakley G. R.* Safeguarding cryptographic keys // Proc. AFIPS 1979 National Computer Conference. V. 48. N. Y., 1979. P. 313–317.

[2] *Shamir A. How to Share a Secret* // Comm. ACM. V. 22, No 1, 1979. P. 612–613.

[3] *Brickell E. F., Davenport D. M.* On the classification of Ideal Secret Sharing Schemes. // J. Cryptology. V. 4, 1991. P. 123–134.

[4] *Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А. Теория кодов, исправляющих ошибки.* М.: Связь, 1979.

[5] *Csirmaz L.* The size of a share must be large // J. Cryptology. V. 10, No 4, 1997. P. 223–232.

[6] *Welsh D. J. A.* Matroid Theory. Academic Press, 1976.

[7] *Блейкли Г. Р., Кабатянский Г. А.* Обобщённые идеальные схемы, разделяющие секрет, и матроиды // Проблемы передачи информации. Т. 33, вып. 3, 1997. С. 102–110.

[8] *Seymour P. O.* On Secret-Sharing Matroids. // J. Comb. Theory. Ser. B. V. 56, 1992. P. 69–73.

[9] *Ashihmin A., Simonis J.* Almost Affine codes. // Designs, codes and cryptography. (В печати.)

По-новому о старом: фрагменты классической математики

Счастливые билеты

С. К. Ландо

Статья составлена по материалам готовящейся к изданию книги «Лекции о производящих функциях», изд-во «Фазис» (лекции были прочитаны в Независимом Московском Университете). Советуем обратиться к этой книге за дополнительными примерами использования производящих функций для решения комбинаторных задач.

Рассмотрим одну популярную в начале 70-х годов задачу, ею как-то А. А. Кириллов открывал свой семинар для младшекурсников. В те времена человек, едущий в городском транспорте в Москве, должен был купить билет в автоматической кассе или у кондуктора. На билетах стояли шестизначные номера.

Билет назывался *счастливым*, если сумма первых трёх цифр его номера равнялась сумме последних трёх цифр.

Так, билеты с номерами 000000 и 123060 — счастливые, а билет с номером 123456 — несчастливый. Считалось, что счастливый билет приносит счастье (особенно, если его съесть).

Возникает вопрос, *сколько всего существует счастливых билетов*? Или: какова вероятность покупки счастливого билета?

Человеку, владеющему элементарными навыками программирования, нетрудно написать программу для подсчёта числа счастливых билетов. Простейшая такая программа перебирает все номера от 000000 до 999999, отбирая среди них счастливые. Давайте, однако, попробуем обойтись без машины.

Разобьем все счастливые билеты на классы, в каждом из которых сумма первых трёх цифр одинакова. Эта сумма может принимать значения от 0 (для тройки цифр 000) до 27 (для тройки 999). Поэтому число классов равно 28. Обозначим через a_n число различных троек цифр с суммой цифр n. Первые несколько значений a_n нетрудно вычислить:

- $a_0 = 1$ (есть всего одна тройка цифр 000 с суммой 0);
- $a_1 = 3$ (есть три тройки 001, 010, 001 с суммой цифр 1);
- $a_2 = 6$ (тройки 002, 020, 200, 011, 101, 110).

Легко видеть, что число счастливых билетов, сумма первых трёх цифр которых равна n, есть a_n^2. Действительно, как в начале, так и в конце номера счастливого билета можно поставить любую тройку цифр с суммой n. Таким образом, для подсчёта числа счастливых билетов нам достаточно вычислить числа a_n, а затем найти сумму квадратов этих 28 чисел.

Для вычисления значений a_n попробуем подсчитать сначала число одно- и двузначных чисел с суммой цифр n. Для каждого $n = 0, 1, 2, \ldots, 9$ существует ровно одно однозначное число с суммой цифр n (запись этого числа совпадает с записью числа n). Будем описывать однозначные числа многочленом

$$A_1(s) = 1 + s + s^2 + \cdots + s^9.$$

Смысл у этого многочлена следующий:

коэффициент при s^n в многочлене A_1 равен числу однозначных чисел, сумма цифр которых равна n.

Другими словами, коэффициент при s^n в многочлене A_1 равен 1, если $0 \leqslant n \leqslant 9$, и равен 0, если $n > 9$.

Выпишем теперь многочлен $A_2(s)$, описывающий двузначные числа. Коэффициент при s^n в многочлене $A_2(s)$ равен числу двузначных чисел с суммой цифр n. (Мы рассматриваем и такие двузначные числа, в которых первая цифра или даже обе цифры могут равняться нулю.)

Нетрудно видеть, что степень многочлена A_2 равна 18. Действительно, 18 — наибольшая возможная сумма цифр двузначного числа. Несложно сосчитать и первые несколько коэффициентов этого многочлена:

$$A_2(s) = 1 + 2s + 3s^2 + 4s^3 + \ldots.$$

Оказывается, многочлен A_2 легко строится по многочлену A_1.

Предложение 1. $A_2(s) = (A_1(s))^2$.

Доказательство. Произведение мономов s^k и s^m даёт вклад в коэффициент при мономе s^n многочлена $(A_1(s))^2$ в том и только в том случае, если $n = k + m$. Поэтому коэффициент при мономе s^n в многочлене $(A_1(s))^2$ есть в точности число способов представить число n в виде суммы $n = k + m, \quad k, m = 0, 1, \ldots, 9$. Таким образом, многочлен в правой части равенства совпадает с многочленом A_2.

Теперь нетрудно выписать и многочлен $A_3(s) = a_0 + a_1 s + \cdots + a_{27} s^{27}$.

Предложение 2. $A_3(s) = (A_1(s))^3$.

Доказательство. Доказательство практически дословно совпадает с доказательством предыдущего утверждения: коэффициент при s^n в многочлене $(A_1(s))^3$ равен числу представлений числа n в виде суммы трёх цифр $n = m + k + l, \quad m, k, l = 0, 1, \ldots, 9$.

Итак, задача о числе счастливых билетов свелась к следующему: надо подсчитать число p_0 — сумму квадратов коэффициентов многочлена $(A_1(s))^3$.

Обратите внимание на то, что умножение на многочлен $A_1(s)$ — очень простая операция. Вычисления можно провести вручную, затратив на них около десяти минут. Надобность в написании программы отпадает.

Однако можно не останавливаться на достигнутом и пойти дальше. Подставим вместо s выражение $e^{i\varphi}$. Тогда $A_3(e^{i\varphi}) = \left(A_1(e^{i\varphi})\right)^3$ будет тригонометрическим полиномом 27-й степени:

$$A_3(e^{i\varphi}) = a_0 + a_1 e^{i\varphi} + \cdots + a_{27} e^{27i\varphi}.$$

Воспользовавшись тем, что

$$\frac{1}{2\pi} \int_0^{2\pi} e^{ik\varphi} \cdot e^{-im\varphi}\, d\varphi = \begin{cases} 1, & k = m, \\ 0, & k \neq m, \end{cases}$$

получим

$$\frac{1}{2\pi} \int_0^{2\pi} |A_3(e^{i\varphi})|^2\, d\varphi =$$

$$\frac{1}{2\pi} \int_0^{2\pi} \left(a_0 + a_1 e^{i\varphi} + \cdots + a_{27} e^{27i\varphi}\right) \left(a_0 + a_1 e^{-i\varphi} + \cdots + a_{27} e^{-27i\varphi}\right) d\varphi =$$

$$= a_0^2 + a_1^2 + \cdots + a_{27}^2.$$

Суммируя геометрическую прогрессию и пользуясь тем, что

$$\frac{e^{i\varphi} - e^{-i\varphi}}{2i} = \sin\varphi,$$

получаем

$$A_1(e^{i\varphi}) = 1 + e^{i\varphi} + \cdots + e^{9i\varphi} = \frac{1 - e^{10i\varphi}}{1 - e^{i\varphi}} = \frac{e^{5i\varphi}}{e^{i\varphi/2}} \frac{\sin 5\varphi}{\sin \frac{\varphi}{2}},$$

откуда искомая величина равна

$$p_0 = \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{\sin^2 5\varphi}{\sin^2 \frac{\varphi}{2}} \right)^3 d\varphi = \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \left(\frac{\sin 10\varphi}{\sin \varphi} \right)^6 d\varphi. \qquad (1)$$

Попробуем оценить значение интеграла (1). График функции $f(\varphi) = \sin(10\varphi)/\sin\varphi$ на отрезке $\left[-\frac{\pi}{2}; \frac{\pi}{2}\right]$ выглядит так, как показано на рис. 1. В нуле функция достигает своего максимума, равного 10. Вне отрезка

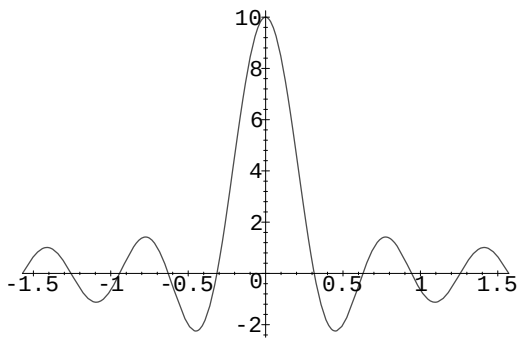


Рис. 1. *Вид графика функции* $f(\varphi) = \frac{\sin(10\varphi)}{\sin\varphi}$

$\left[-\frac{\pi}{10}; \frac{\pi}{10}\right]$ величина функции f не превосходит $\frac{1}{\sin\frac{\pi}{10}} \approx 3$. Поэтому вклад дополнения к отрезку $\left[-\frac{\pi}{10}; \frac{\pi}{10}\right]$ в интеграл (1) не превосходит $\pi \cdot 3^6 \approx 2100$ (на самом деле он значительно меньше).

Основная же составляющая интеграла (1) сосредоточена на отрезке $\left[-\frac{\pi}{10}; \frac{\pi}{10}\right]$. Для оценки вклада этого отрезка воспользуемся *методом стационарной фазы*. Этот метод позволяет оценить значение интеграла

$$\int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} f^t d\varphi = \int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{t \ln f} d\varphi$$

при $t \to \infty$. При больших значениях t величина интеграла определяется поведением функции $\ln f$ («фазы») в окрестности своей стационарной

точки 0 (точки, в которой $(\ln f)' = 0$, или, что то же самое, $f' = 0$). В окрестности нуля $f(\varphi) \approx 10\left(1 - \frac{33}{2}\varphi^2\right)$, а $\ln f(\varphi) \approx \ln 10 - \frac{33}{2}\varphi^2$. При больших t имеем

$$\int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{t\left(\ln 10 - \frac{33}{2}\varphi^2\right)} d\varphi = e^{t\ln 10} \int_{-\frac{\pi}{10}}^{\frac{\pi}{10}} e^{-\frac{33}{2}t\varphi^2} d\varphi \approx e^{t\ln 10} \frac{\sqrt{2\pi}}{\sqrt{33t}}$$

Полагая $t = 6$ и вспоминая формулу (1), получаем приближённое равенство

$$p_0 \approx \frac{10^6}{3\sqrt{11\pi}} \approx 56700.$$

Полученный результат с хорошей точностью (отклонение составляет не более 3%) приближает искомое значение[1)].

Некоторые итоги

На основании рассмотренного примера можно сделать некоторые выводы о комбинаторных задачах и методах их решения.

Задачи перечислительной комбинаторики состоят в подсчёте числа объектов, принадлежащих некоторому семейству конечных множеств. У каждого множества семейства имеется свой номер (в задаче о числе счастливых билетов таким номером была сумма цифр трёхзначного числа).

Как правило, задача перечислительной комбинаторики «в принципе» разрешима: для каждого множества из семейства можно выписать все его элементы и таким образом узнать их число. Проблема, однако, состоит в том, чтобы найти «хорошее» решение, не требующее выписывания всех элементов изучаемых множеств.

Определить, что такое хорошее решение, довольно трудно. Зачастую можно лишь сравнить два решения и сказать, какое из них лучше.

При решении задач перечислительной комбинаторики очень полезно рассматривать производящие многочлены (или, более общо, производящие ряды). В нашем случае пользу принес производящий многочлен A_3. Операции с комбинаторными объектами очень естественно выражаются в терминах производящих функций. Так, переход от однозначных чисел с заданной суммой цифр к трёхзначным числам состоял просто в возведении производящего многочлена A_1 в куб.

[1)] Прим. ред.: Как-то участникам Всесоюзной математической олимпиады во время отдыха кто-то предложил задачу о счастливых билетах. Для всех ребят задача была новой. Большинство стало активно решать её. Прошло некоторое время, и один из школьников изложил решение, приведённое выше, включая формулу (1) и оценку числа счастливых билетов. Это был будущий филдсовский лауреат Владимир Дринфельд. (О филдсовских медалях см. статьи, помещённые в этом сборнике, стр. 19–40.)

Привлечение методов из смежных областей математики (например, из анализа) позволяет по-иному взглянуть на перечислительную задачу и найти новые, зачастую неожиданные, подходы к её решению.

Задачи

В заключение предлагаем несколько задач.

1. Докажите, что счастливых билетов ровно 55 252 штуки. (Используйте любой из обсуждавшихся способов или придумайте свой.)

2. Докажите, что число счастливых билетов равно

$$\binom{32}{5} - \binom{6}{1}\binom{22}{5} + \binom{6}{2}\binom{12}{5}.$$

3. Найдите выражение для числа счастливых билетов из $2r$ цифр в системе счисления с основанием q.

Список литературы

[1] *Савин А. П., Финк Л. М.* Разговор в трамвае. // Квант. 1975. №7. С. 67–70.

[2] *Финк Л. М.* Ещё раз о счастливых билетах // Квант. 1976. №12. С. 68–70.

[3] *Полиа Г., Сеге Г.* Задачи и теоремы из анализа. Т. 1. М.: Наука, 1978. Задача №30.

Арифметический минимум квадратичной формы и сферические коды

Н. Н. Андреев В. А. Юдин

Посвящается 150-летию со дня рождения Е. И. Золотарева (1847–1878).

В этой статье мы расскажем об одной старой задаче геометрии чисел: *Как много точек с целыми координатами могут располагаться на поверхности эллипсоида в d-мерном евклидовом пространстве, если внутри него нет точек с целыми координатами, за исключением его центра — нуля?* Сформулируем задачу более точно.

Пусть $\mathbb{Z}^d$ — решётка целых чисел в $\mathbb{R}^d$, т. е. множество точек, у которых все координаты — целые числа; точку $x \in \mathbb{Z}^d$ будем называть целой точкой. Обозначим через $xy = x_1y_1 + \cdots + x_dy_d$ скалярное произведение векторов $x = (x_1, \ldots, x_d)$ и $y = (y_1, \ldots, y_d)$, а через $|x| = \sqrt{xx}$ — норму вектора x. Пусть $Axx = \sum_{i,j=1}^{d} a_{ij}x_ix_j$ — положительно определённая квадратичная форма (т. е. $Axx > 0$ для всех $0 \neq x \in \mathbb{R}^d$), порожденная матрицей $A = (a_{ij})_{i,j=1}^{d}$ с определителем $D(A)$. Число

$$\gamma(A) = \inf_{x \in \mathbb{Z}^d \setminus \{0\}} Axx$$

называется арифметическим минимумом квадратичной формы, а величина

$$N(d, A) = \sum_{x \in \mathbb{Z}^d \,:\, Axx = \gamma(A)} 1$$

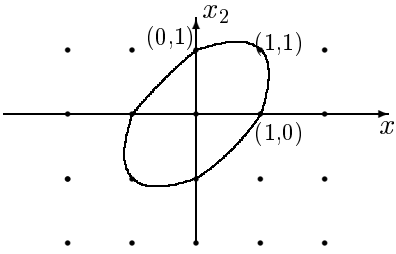
называется числом представлений ее минимума. Можно показать, что положительная определённость формы Axx гарантирует достижение инфимума $\gamma(A)$.

Используя введенные обозначения, интересующую нас задачу можно сформулировать следующим образом: требуется найти числа

$$N_d = \sup_A N(d, A),$$

где супремум берётся по множеству всех положительно определённых матриц A.

Так, при $d = 2$ требуется расположить эллипс с центром в начале координат так, чтобы внутри него целых точек не было, а на границе их количество стало максимальным. Докажем, что $N_2 = 6$.



Это значение достигается на квадратичной форме

$$Axx = x_1^2 - x_1x_2 + x_2^2.$$

Арифметический минимум $\gamma(A) = 1$ достигается в шести точках $\pm(1,0)$, $\pm(0,1)$, $\pm(1,1)$.

Покажем теперь, что $N_2 \leqslant 6$. Ввиду симметрии эллипса относительно начала координат для каждой точки эллипса противоположная ей также принадлежит эллипсу. Пусть $\pm x^1, \ldots, \pm x^q$ — все целые точки, на которых достигается минимальное значение формы A (*минимальные целые точки*). Выберем из каждой пары противоположных минимальных точек по одной, обозначим их $x^1, \ldots, x^q$ и разобьем на 4 класса: к классу A_0 отнесем точки, у которых обе координаты — чётные; к классу A_1 — точки, у которых первая координата нечётная, а вторая чётная; класс A_2 будет состоять из точек, у которых первая координата чётная, а вторая нечётная; класс A_3 — из точек, у которых обе координаты нечётные. Каждый класс A_i, $i = 0, 1, 2, 3$ состоит не более, чем из одной минимальной точки. Действительно, допустим, что какой-то из классов содержит не меньше двух точек, например $x, y \in A_i$, тогда их полусумма $\frac{x+y}{2}$ также будет отличной от нуля целой точкой. Ввиду строгой выпуклости эллипса она будет расположена внутри него, что невозможно, так как по условию внутри эллипса нет целых точек, кроме нуля. В классе A_0 вообще нет ни одной минимальной точки (если $x \in \mathbb{Z}^2$, $x \neq 0$, $x \in A_0$, то $\frac{1}{2}x \in \mathbb{Z}^2$, чего не может быть). Таким образом, $q \leqslant 3$ или $N_2 \leqslant 6$. Приведенный выше пример показывает достижимость полученной оценки.

Г. Ф. Вороной провёл аналогичное рассуждение в $\mathbb{R}^d$ и получил оценку $N_d \leqslant 2(2^d - 1)$, однако при $d \geqslant 3$ она оказывается грубой. Так, при $d = 3$ получается оценка $N_3 \leqslant 14$, а на самом деле $N_3 = 12$, как будет показано ниже. Из работы А. Н. Коркина и Е. И. Золотарева [2] непосредственно вытекает оценка снизу $N_d \geqslant d(d+1)$. (Результаты Г. Ф. Вороного, А. Н. Коркина и Е. И. Золотарева приведены в [3].)

Что касается точных значений числа N_d, то при $d \leqslant 6$ их можно получить из работы [2] и из работы Барнса 1957 года. Было доказано, что квадратичные формы, на которых достигается максимум N_d, следует искать среди так называемых предельных квадратичных форм, т.е. таких,

для которых постоянная Эрмита

$$\gamma_d = \sup_A \frac{\gamma(A)}{\sqrt[d]{D(A)}}$$

принимает наибольшее возможное значение. Задача вычисления чисел γ_d непосредственно связана с вопросом наиплотнейшей решётчатой упаковки шаров в евклидовом пространстве и с задачей, рассматриваемой в этой статье, как мы увидим позже. Исследования начинаются работами Гаусса и Лагранжа, которые показали, что $\gamma_2 = 2/\sqrt{3}$, $\gamma_3 = \sqrt[3]{2}$. В 1872 году А. Н. Коркин и Е. И. Золотарев в их первой совместной работе [1] доказали, что $\gamma_4 = \sqrt{2}$, а в работе 1877 года вычислили $\gamma_5 = \sqrt[5]{8}$. К настоящему времени найдены точные значения γ_d до $d = 8$ (в двух работах 1925 и 1935 годов Блихфельд показал, что $\gamma_6 = \sqrt[6]{64/3}$, $\gamma_7 = \sqrt[7]{64}$, $\gamma_8 = 2$). Крайне интересным является вопрос о поведении γ_d при $d \to \infty$. Подробно полученные результаты изложены в [4].

В [5] Ватсон упростил способы нахождения чисел N_d для $d \leqslant 6$ и нашёл точные значения для $d = 7, 8, 9$, проводя достаточно длинные вычисления с квадратичными формами. В конце 70-х годов В. И. Левенштейном и независимо Н. Слоэном и А. Одлыжко было показано, что $N_8 = 240$ и $N_{24} = 196560$. После этих работ таблица известных точных значений N_d приняла вид

d	2	3	4	5	6	7	8	9	24
N_d	6	12	24	40	72	126	240	272	196560

В остальных размерностях точные значения не известны.

Приведем аналитический способ получения оценок сверху чисел N_d. Вначале дадим несколько определений.

Пусть $e_1, \ldots, e_d$ — линейно-независимая система векторов из $\mathbb{R}^d$. Решеткой называется множество

$$L = \{k_1 e_1 + k_2 e_2 + \cdots + k_d e_d\}_{k=(k_1,k_2,\ldots,k_d)\in\mathbb{Z}^d}.$$

Вектора, для которых

$$|x| = \inf_{x\in L\setminus\{0\}} |x| \stackrel{\text{def}}{=} \mu_1(L),$$

называются минимальными векторами решётки, совокупность этих векторов обозначим V. Очевидно, что для каждого вектора $x \in V$ противоположный вектор $-x$ также лежит в V. Теперь нетрудно заметить,

что для любых двух различных векторов $x, y \in V$ выполнено неравенство $xy \leqslant 1/2\mu_1^2(L)$, т. е. угол между ними не менее 60°. Без ограничения общности будем считать $\mu_1(L) = 1$.

Возвращаясь к нашей задаче, сделаем линейную замену переменных $x = Ty$, приводящую исходную квадратичную форму к виду $y_1^2 + \cdots + y_d^2$. Тем самым вопрос свёлся к поиску в d-мерном евклидовом пространстве решётки L, у которой максимальное количество минимальных векторов.

Вследствие перечисленных выше свойств множества V эта задача оказывается тесно связанной с двумя другими интересными задачами.

Какое максимальное количество точек B_d может иметь симметричный сферический $1/2$-код в размерности d? Другими словами, *какое максимальное количество точек можно разместить на единичной сфере S^{d-1} с условиями, что каждая точка имеет противоположную и что модуль скалярного произведения любых двух различных из них и не противоположных не превосходит $1/2$?*

Найти контактное число M_d шаров в размерности d, т. е. максимальное количество шаров одинакового радиуса, которые могут касаться одного данного шара. Иначе говоря — *какое максимальное количество точек можно расположить на единичной сфере так, чтобы скалярное произведение любых двух из них не превосходило $1/2$?*

Точное значение M_d до недавнего времени было известно только при $d = 2$ ($M_2 = 6$) и $d = 3$ ($M_3 = 12$). Наилучшие известные оценки этой величины в зависимости от размерности приведены в [6, т.1, с.42].

Понятно, что

$$N_d \leqslant B_d \leqslant M_d. \tag{1}$$

Мы будем оценивать сверху величину B_d и тем самым получать оценку сверху для количества минимальных векторов в решетке размерности d. При этом мы воспользуемся идеями, предложенными в 1968 году П. Дельсартом. Именно он стал использовать в геометрических задачах положительную определённость. Итак, через $\{P_k^d(t)\}_{k=1}^{\infty}$ обозначим систему многочленов Гегенбауэра с нормировкой $P_k^d(1) = 1$:

$$P_0^d(t) = 1, \quad P_1^d(t) = t, \quad P_2^d(t) = \frac{dt^2 - 1}{d-1}, \quad P_3^d(t) = \frac{(d+2)t^3 - 3t}{d-1}, \ldots$$

$$(k + d - 2)P_{k+1}^d(t) = (2k + d - 2)tP_k^d(t) - kP_{k-1}^d(t).$$

На интервале $(-1; 1)$ они образуют ортогональную систему многочленов

с весом $(1-t^2)^{\frac{d-3}{2}}$, т. е.

$$\int_{-1}^{1} P_k^d(t)P_l^d(t)(1-t^2)^{\frac{d-3}{2}} = 0, \quad k \neq l.$$

В геометрических задачах используется их положительная определённость: для любого конечного множества точек $x^{(1)}, \ldots, x^{(N)}$ из S^{d-1}, любого $s \in \mathbb{N}$ и любых $p_k,\ p_l \in \mathbb{R}$ справедливо неравенство

$$\sum_{k,l=1}^{N} P_s^d(x^{(k)}x^{(l)})p_k p_l \geqslant 0. \tag{2}$$

Пусть в d-мерном евклидовом пространстве нам дан симметричный сферический 1/2-код, состоящий из N точек, т. е. множество$\{x^{(i)}\}_{i=1}^{N}$ точек в S^{d-1} таких, что $-1/2 \leqslant x^{(i)}x^{(j)} \leqslant 1/2$ при $x^{(i)} \neq \pm x^{(j)}$. Рассмотрим непрерывную на отрезке $[-1;1]$ функцию $h(t)$ такую, что $h(t) \leqslant 0$ при $t \in [-1/2;1/2]$ и все ее коэффициенты Фурье по системе многочленов Гегенбауэра неотрицательны:

$$h(t) = \sum_{k=0}^{\infty} \widehat{h}_k P_k^d(t), \quad \widehat{h}_k \geqslant 0,\ \widehat{h}_0 > 0.$$

В такой ситуации имеем:

$$I = \sum_{k,l=1}^{N} h(x^{(k)}x^{(l)}) = \sum_{s=0}^{\infty} \widehat{h}_s \sum_{k,l=1}^{N} P_s^d(x^{(k)}x^{(l)}) \geqslant \widehat{h}_0 \sum_{k,l=1}^{N} 1 = N^2\widehat{h}_0,$$

где неравенство верно вследствие неотрицательности коэффициентов $\widehat{h}_s$ и положительной определённости многочленов Гегенбауэра. С другой стороны, так как $h(x^{(k)}x^{(l)}) \leqslant 0$ при $x^{(k)} \neq \pm x^{(l)}$, то

$$\begin{aligned} I = \sum_{k,l=1}^{N} h(x^{(k)}x^{(l)}) = \\ = \sum_{x^{(k)}=x^{(l)}} h(x^{(k)}x^{(l)}) + \sum_{x^{(k)}=-x^{(l)}} h(x^{(k)}x^{(l)}) + \sum_{x^{(k)}\neq\pm x^{(l)}} h(x^{(k)}x^{(l)}) \leqslant \\ \leqslant N(h(1)+h(-1)). \end{aligned}$$

Таким образом

$$N \leqslant \frac{h(1)+h(-1)}{\widehat{h}_0}. \tag{3}$$

Подбирая нужным образом функцию $h(t)$, получаем оценки сверху. Заметим, что число N натуральное, поэтому оно на самом деле оценивается целой частью выражения, стоящего в правой части неравенства (3). Подобным способом получены точные оценки сверху для чисел M_8 и M_{24} и тем самым для чисел N_8 и N_{24}, а решётки с соответствующим количеством минимальных векторов были уже известны: в случае $d = 8$ экстремальной оказалась решётка Коркина-Золотарева E_8, а в случае $d = 24$ — решётка Лича.

Используем неравенство (3) для оценки мощности симметрического 1/2-кода. Рассмотрим многочлен

$$h(t) = t^2(t^2 - \frac{1}{4}).$$

Напишем его разложение в ряд Фурье:

$$h(t) = \frac{d^2-1}{(d+2)(d+4)}P_4^d(t) + \frac{(20-d)(d-1)}{4d(d+4)}P_2^d(t) + \frac{10-d}{4d(d+2)}P_0^d(t). \quad (4)$$

Все его коэффициенты положительны при $2 \leqslant d \leqslant 9$. Следовательно, из неравенства (3) найдем

$$N_d \leqslant \frac{6d(d+2)}{10-d}.$$

Эта оценка оказывается точной в размерностях $d = 3, 4, 6, 7, 8$. Аналогичные (4) многочлены можно строить и в высших размерностях. Рассматривая многочлены

$$h(t) = t^4(t^2 - \frac{1}{4}), \quad h(t) = t^6(t^2 - \frac{1}{4}), \quad h(t) = t^2(t^2 - \frac{1}{16})^2(t^2 - \frac{1}{4}),$$

получаем следующие оценки сверху мощности симметрического 1/2-кода и вследствие неравенства (1) — числа минимальных векторов в решётке:

d	3	4	5	6	7	8	9	10	11	12	13
$N_d \leqslant$	12	24	42	72	126	240	367	560	858	1344	2210

d	17	18	19	20	21	22	23	24
$N_d \leqslant$	11683	16298	22866	32445	46947	70200	111136	196560

Приведем примеры наилучших квадратичных форм и решёток, для которых оценки, представленные в таблице, при $d \leqslant 8$ достигаются. Для

$d = 2$ ранее уже был дан очевидный пример квадратичной формы $x_1^2 \pm$ $\pm x_1x_2 + x_2^2$. Для $d = 3$ квадратичная форма имеет вид $Axx = x_1^2 + x_2^2 +$ $+ x_3^2 + x_1x_2 + x_1x_3 + x_2x_3$. Легко устанавливается (выделением полных квадратов), что она положительно определена.

Кроме того, она целочисленна: $Axx \in \mathbb{Z}$ для любого $x \in \mathbb{Z}^3$. Значит, ее арифметический минимум не меньше 1, т. е. $\min_{x\in\mathbb{Z}^3\setminus\{0\}} Axx \geqslant 1$. На самом деле он равен 1 и достигается в 12 точках $\pm(1,0,0)$, $\pm(0,1,0)$, $\pm(0,0,1)$, $\pm(1,-1,0)$, $\pm(1,0,-1)$, $\pm(0,1,-1)$.

Эту последовательность можно продолжить, но нам удобнее иметь дело с решётками (что по смыслу одно и то же). Это даст нам возможность заметить одно весьма любопытное явление: «сечения» минимальных векторов решётки Коркина-Золотарева из $\mathbb{R}^8$ дают экстремальные конструкции для всех меньших размерностей.

Минимальные вектора решётки Коркина-Золотарева E_8 состоят из двух групп векторов. В первую входят 128 векторов вида $\frac{1}{\sqrt{8}}(\pm1, \pm1, \ldots,$ $\pm1)$ с чётным числом (0,2,4,6,8) плюсов. Во вторую входят вектора, у которых две произвольные координаты равны $\pm1/\sqrt{2}$, а остальные — 0. Их количество $C_8^2 \cdot 4 = 112$. В сумме получим 240 векторов. Их совокупность (обозначим ее через W_8) и есть экстремальный набор векторов при $d = 8$.

Замечая, что «сечение» решётки гиперплоскостью $ax = 0$ снова есть решётка, а минимальные вектора первоначальной решётки становятся минимальными векторами «сечения», возьмём сечение в $\mathbb{R}^8$ решётки Коркина-Золотарева гиперплоскостью $ax = 0$, $a = (1, 1, \ldots, 1)$ и положим

$$W_7 = \{x \in W_8 \ : \ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 0\}.$$

Подсчитаем количество элементов в W_7. Из первой группы минимальных векторов решётки Коркина-Золотарева в W_7 войдут лишь те вектора, у которых 4 координаты равны $+1$ и четыре равны -1. Их количество равно $C_8^4 = 70$. Из второй группы векторов лишь половина (координаты имеют разные знаки) удовлетворяет уравнению $ax = 0$. Таким образом, W_7 содержит $70 + 56 = 126$ векторов и является экстремальным набором векторов при $d = 7$.

Аналогично можно доказать, что множества

$$W_6 = \{x \in W_8 \ : \ x_1 + x_2 = 0, \ x_3 + x_4 + x_5 + x_6 + x_7 + x_8 = 0\},$$
$$W_5 = \{x \in W_8 \ : \ x_1 + x_2 = 0, \ x_3 + x_4 + x_5 + x_6 = 0, \ x_7 + x_8 = 0\},$$
$$W_4 = \{x \in W_8 \ : \ x_1 + x_2 = 0, \ x_3 + x_4 = 0, \ x_5 + x_6 = 0, \ x_7 + x_8 = 0\},$$

являясь наборами минимальных векторов решёток в $\mathbb{R}^6$, $\mathbb{R}^5$, $\mathbb{R}^4$, содержат соответственно 72, 40 и 24 вектора.

В заключение отметим, почему предложенный метод получения оценок сверху чисел N_d «грубит», например, при $d = 5$. Полученная оценка $N_5 \leqslant 42$ не является точной — на самом деле $N_5 = 40$. Чтобы ответить на этот вопрос, надо проанализировать предложенное нами доказательство. При $d = 5$ оценка оказывается точной в тех случаях, когда неравенство (2) обращается в равенство для экстремальной конструкции при $s = 1, 2, 3, 4$. Однако этого не происходит для W_5: $\sum_{x,y \in W_5} P_4^5(xy) > 0$. Аналогичная ситуация и при $d = 3$, но тут «везёт» в вычислениях: $[\frac{90}{7}] = 12$.

Итак, для размерностей $d = 3,\ 4,\ 6,\ 7,\ 8,\ 24$ мы привели, как нам кажется, наиболее простой метод нахождения чисел N_d. При $d = 10,\ 11,\ 12,\ 13,\ 17,\ 18,\ 19,\ 20,\ 21,\ 22,\ 23$ получены новые оценки сверху для количества минимальных векторов решёток в этих размерностях.

Список литературы

[1] *Korkine A., Zolotareff G.* Sur les formes quadratiques positives quaternaires // Math. Ann. V. 5. 1872. P. 581–583.
Русск. пер.: *Золотарев Е. И.* Полное собр. соч., вып. 1. Изд. АН СССР, 1931.

[2] *Korkine A., Zolotareff G.* Sur les formes quadratiques positives // Math. Ann. V.11. 1877. P. 242–292.
Русск. пер.: там же.

[3] *Делоне Б. Н.* Петербургская школа теории чисел. М.-Л.: Изд. АН СССР, 1947.

[4] *Рышков С. С., Барановский Е. П.* Классические методы теории решетчатых упаковок // УМН. Т. 34, вып. 4. 1974. С. 3–63.

[5] *Watson G. L.* The number of minimum points of a positive quadric form // Dissertationes mathematicae. LXXXIV. 1971, pp. 2–42.

[6] *Конвей Дж., Слоэн Н.* Упаковки шаров, решетки и группы. М.: Мир, 1990.

Решение обобщённой задачи Мальфатти с помощью комплексной (гиперболической) тригонометрии

В. З. Беленький А. А. Заславский

Как-то во время летнего отдыха мы задумались над следующей задачей: в углы данного $\triangle ABC$ вписать окружности так, чтобы они попарно касались друг друга. Несмотря на классически простую формулировку, тогда нам не удалось найти в литературе упоминаний об этой задаче и мы решили её самостоятельно. Впоследствии мы узнали[1)], что эта задача давно известна как задача итальянского математика Мальфатти, опубликованная им в 1803 году [1], и её решению посвящена обширная литература.

Сам Мальфатти[2)] в своей публикации приводит конечные формулы для искомых радиусов, но не даёт их вывода, указав лишь, что алгебраические выкладки очень громоздки. В 1826 году чисто геометрическое решение, и тоже без доказательства, предложил Я. Штейнер [2] — один из крупнейших геометров прошлого века. Полное доказательство метода Штейнера впервые было дано Шретером в 1874 году, его изложение можно найти в прекрасном сборнике задач на геометрические построения Ю. Петерсена [3]. Все эти сведения приводятся в книге А. Адлера, изданной в русском переводе в Одессе в 1910 году [4]. Прямое геометрическое решение задачи даётся в одной из книг серии «Библиотека математического кружка», см. [5].

Решение, найденное нами, оказалось оригинальным по подходу и совершенно отличным от методов, использовавшихся ранее. Нам кажется, что этот метод представляет самостоятельный интерес, причем не только для первоначальной геометрической задачи. Предложенное нами решение было опубликовано в журнале «Квант» [6].

[1)] Авторы признательны А. А. Егорову и А. А. Фридману за библиографические сведения.

[2)] Как сообщил В. В. Прасолов, Мальфатти решал другую задачу: поместить в треугольник три непересекающиеся круга максимальной общей площади. Легко убедиться, что решение, приведённое Мальфатти, не всегда оптимально. Более удивителен тот факт, что оно ***никогда не оптимально***.

В учебнике Ж. Адамара [7] задача Мальфатти рассмотрена в обобщённой постановке и дано её полное решение по методу Штейнера – Шретера – Петерсена. Здесь мы дадим решение обобщённой задачи Мальфатти, развивая основную идею нашего метода.

Обобщённая задача Мальфатти. *Даны три прямые общего положения. Построить три взаимно касающиеся окружности так, чтобы окружности O_1 и O_2 касались прямой l_3, окружности O_2 и O_3 касались прямой l_1, окружности O_3 и O_1 касались прямой l_2.*

Имеется 10 различных конфигураций, порождающих для данной тройки прямых 32 решения (см. рис. 1). Далее мы получим единообразным способом формулы для радиусов искомых окружностей во всех 32 случаях.

1. Основная система уравнений

Решения задачи Мальфатти будут получаться из решений систем вида

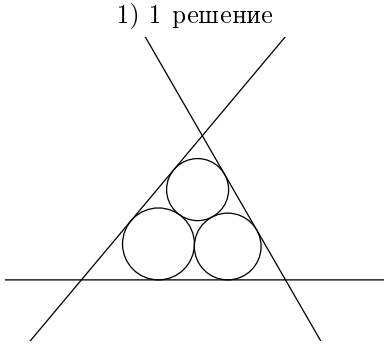
$$\begin{cases} u^2+ & 2uv\cdot\sqrt{1-c/p}+ & v^2 =c, \\ v^2+ & 2vw\cdot\sqrt{1-a/p}+ & w^2 =a, \\ w^2+ & 2wu\cdot\sqrt{1-b/p}+ & v^2 =b, \end{cases} \tag{1}$$

где неизвестные u, v, w — комплексные числа (как будет видно далее, в решениях задачи Мальфатти эти неизвестные будут либо вещественными, либо чисто мнимыми), а (комплексные) параметры a, b, c, p удовлетворяют условиям

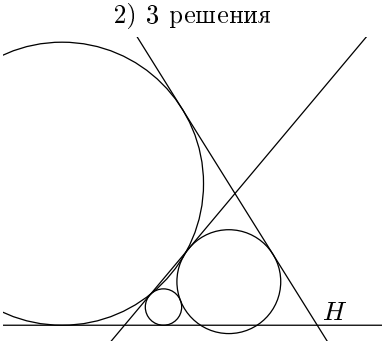
$$p=\frac{a+b+c}{2}, \quad a\neq 0,\ b\neq 0,\ c\neq 0, \quad D\stackrel{\text{def}}{=}p(p-a)(p-b)(p-c)\neq 0.$$

Ниже (см. п. 3, стр. 145) будет получено решение системы (1), исходящее из геометрических соображений, когда уравнения в (1) понимаются как утверждения теоремы косинусов для трёх треугольников, вписанных в одну и ту же окружность. Условие равенства сумм углов этих треугольников 180 градусам даёт линейные уравнения для углов, противолежащих сторонам. После решения этих уравнений стороны вычисляются по теореме синусов.

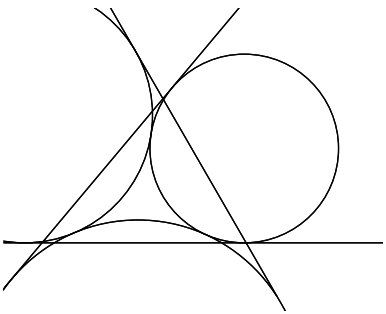
Для решения обобщённой задачи Мальфатти придётся рассматривать такие «треугольники», стороны которых являются произвольными комплексными числами (точнее, как будет показано, потребуются треугольники с отрицательными и чисто мнимыми длинами сторон). Поэтому начнем с краткого объяснения, что такое «треугольник с комплексными сторонами».



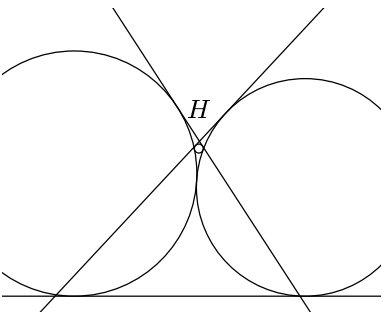
1) 1 решение



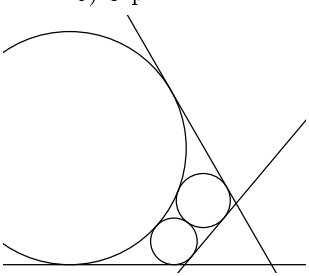
2) 3 решения



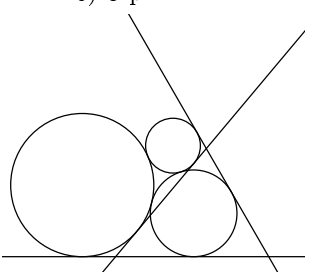
3) 1 решение



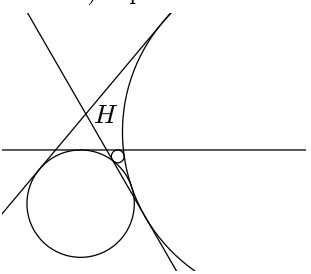
4) 3 решения



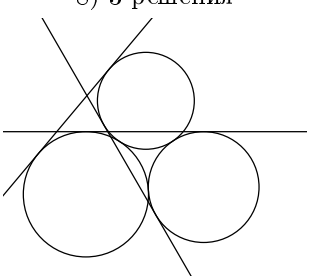
5) 3 решения



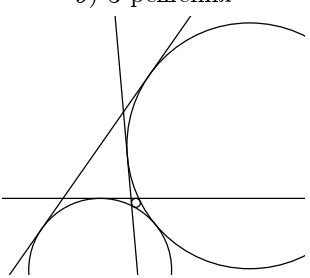
6) 3 решения



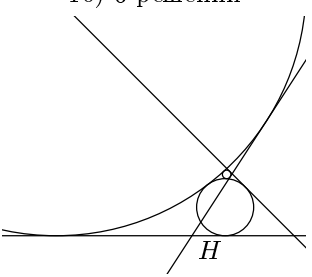
7) 6 решений



8) 3 решения



9) 3 решения



10) 6 решений

Рис. 1. *Конфигурации обобщённой задачи Мальфатти.*

2. Комплексные треугольники

Определение 1. Пусть (a, b, c) — тройка произвольных комплексных чисел, для которой величина

$$D = p(p-a)(p-b)(p-c) = \frac{1}{16}\left(2a^2b^2 + 2b^2c^2 + 2c^2a^2 - a^4 - b^4 - c^4\right) \neq 0,$$

здесь $p \stackrel{\text{def}}{=} (a+b+c)/2$. Будем называть такую тройку *комплексным треугольником*, а числа a, b, c — *сторонами* этого треугольника.

Будем последовательно развивать эту геометрическую аналогию. Величину $S = \pm\sqrt{D}$ естественно назвать *площадью* треугольника (формула Герона). Площадь считаем в комплексном случае определённой с точностью до знака. Величину $R = abc/4S$ назовем *радиусом описанной окружности*.

Определим углы комплексного треугольника. По аналогии с теоремами синусов и косинусов имеем равенства

$$\cos\alpha = \frac{b^2 + c^2 - a^2}{2bc}, \qquad \sin\alpha = \frac{a}{2R}, \tag{2}$$

которые будем считать определением углов треугольника. Прямое вычисление показывает, что

$$\sin^2\alpha + \cos^2\alpha = \left(\frac{b^2 + c^2 - a^2}{2bc}\right)^2 + \frac{4D}{b^2c^2} = 1,$$

поэтому корректность определения углов обеспечивается следующей леммой.

Лемма 1. *Для любых двух комплексных чисел z_1, z_2, удовлетворяющих соотношению*

$$z_1^2 + z_2^2 = 1,$$

найдётся такое комплексное число $\psi = \varphi + it$, $\varphi, t \in \mathbb{R}$, что

$$\cos\psi = z_1, \quad \sin\psi = z_2,$$

причем t определено однозначно, а φ определено с точностью до $2k\pi$, где $k \in \mathbb{Z}$.

Доказательство. Запишем число $z = z_1 + iz_2$ в тригонометрической форме $z = r(\cos\varphi + i\sin\varphi)$, этим число r определено однозначно, а число φ — с точностью до $2k\pi$, $k \in \mathbb{Z}$. Полагаем $t = -\ln r$, $\psi = \varphi + it$.

Далее используем формулу Эйлера $e^{ix} = \cos x + i \sin x$ и вытекающие из неё выражения для синуса $\sin x = \frac{1}{2i}(e^{ix} - e^{-ix})$ и косинуса $\cos x = \frac{1}{2i}(e^{ix} + e^{-ix})$ для того, чтобы вычислить синус и косинус числа ψ:

$$z = e^{-t} \cdot (\cos\varphi + i \sin\varphi) = e^{i\psi}$$
$$\frac{1}{z} = \frac{1}{z_1 + iz_2} = z_1 - iz_2$$
$$\cos\psi = \frac{1}{2} \cdot (z + \frac{1}{z}) = z_1$$
$$\sin\psi = \frac{1}{2i} \cdot (z - \frac{1}{z}) = \frac{1}{2i} \cdot 2iz_2 = z_2.$$

Теорема 1. *Сумма углов комплексного треугольника с точностью до $2k\pi$, $k \in \mathbb{Z}$, равна π.*

Доказательство. Другими словами, $\cos(\alpha + \beta + \gamma) = -1$, где через α, β, γ обозначены углы треугольника. Справедливость этого равенства проверяется вычислением с использованием (1) и тригонометрических формул сложения.

Замечание. Верно и обратное. Любые комплексные числа α, β, γ, для которых выполнено $\alpha + \beta + \gamma = \pi$, являются углами некоторого комплексного треугольника. (Его стороны равны $\lambda \sin\alpha$, $\lambda \sin\beta$, $\lambda \sin\gamma$.)

3. Решение основной системы

Введём новые параметры $\tilde{a} = \pm\sqrt{a}$, $\tilde{b} = \pm\sqrt{b}$, $\tilde{c} = \pm\sqrt{c}$, $\tilde{R} = \pm\sqrt{p}/2$. Тогда уравнения системы (1) есть не что иное, как теорема косинусов, записанная для треугольников $(u, v, \tilde{c})$, $(v, w, \tilde{a})$, $(w, u, \tilde{b})$, у которых радиус описанной окружности равен $\tilde{R}$.

Углы, противолежащие сторонам $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, однозначно определяются через параметры $\tilde{a}$, $\tilde{b}$, $\tilde{c}$ и $\tilde{R}$. Обозначая эти углы через $\hat{a}$, $\hat{b}$, $\hat{c}$, имеем:

$$\sin\hat{a} = \frac{\tilde{a}}{2\tilde{R}}, \quad \sin\hat{b} = \frac{\tilde{b}}{2\tilde{R}}, \quad \sin\hat{c} = \frac{\tilde{b}}{2\tilde{R}},$$
$$\cos\hat{a} = -\sqrt{1 - \frac{a}{p}}, \ \cos\hat{b} = -\sqrt{1 - \frac{b}{p}}, \ \cos\hat{c} = -\sqrt{1 - \frac{c}{p}}.$$

Параметры $\tilde{a}, \tilde{b}, \tilde{c}$ определены с точностью до знака. Перемена знака у одного из них меняет знак соответствующего угла ($\hat{a}$, $\hat{b}$ или $\hat{c}$), а остальные два угла в соответствующем треугольнике ($(v, w, \tilde{a})$, $(w, u, \tilde{b})$ или $(u, v, \tilde{c})$) меняет на смежные ($\hat{x} \mapsto \pi - \hat{x}$) в силу теоремы косинусов.

У остальных углов в треугольниках $(v, w, \tilde{a})$, $(w, u, \tilde{b})$, $(u, v, \tilde{c})$ однозначно определены (через параметры и неизвестные) только синусы, которых, впрочем, достаточно для выражения исходных неизвестных через эти углы. Мы хотим перейти к угловым неизвестным $\hat{u}$, $\hat{v}$, $\hat{w}$, относительно которых будем иметь по теореме 1 линейную систему уравнений (сумма углов в треугольниках равна π). Из-за указанной неоднозначности получаем такой набор систем линейных уравнений

$$\begin{cases} \frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{u})+\frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{v})+\hat{c}=\pi\ (\mathrm{mod}\ 2\pi), \\ \frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{v})+\frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{w})+\hat{a}=\pi\ (\mathrm{mod}\ 2\pi), \\ \frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{w})+\frac{\pi}{2}\pm\ (\frac{\pi}{2}-\hat{u})+\hat{b}=\pi\ (\mathrm{mod}\ 2\pi). \end{cases} \tag{3}$$

Поскольку исходные неизвестные u, v, w не меняются при замене углов $\hat{u}$, $\hat{v}$, $\hat{w}$ на смежные, систем (3) не 64, как кажется на первый взгляд, а всего лишь 8 (важно лишь с одинаковыми или разными знаками входит каждая переменная в уравнения этих систем).

Проанализируем возможные расстановки знаков. Если нечётное число переменных входит с разными знаками в уравнения системы (3), то такая система, как нетрудно видеть, вырожденная, и для её разрешимости должны выполняться соотношения на параметры. Эти соотношения после подходящей перемены знаков у углов $\hat{a}$, $\hat{b}$, $\hat{c}$ приводятся к виду $\hat{a}+\hat{b}+\hat{c}=0$.

Покажем, что условие $\hat{a}+\hat{b}+\hat{c}=0$ влечёт $D=0$. По замечанию к теореме 1 углы $\pi-\hat{a}$, $\pi-\hat{b}$, $\pi-\hat{c}$ являются углами треугольника, стороны которого равны с точностью до пропорциональности $\sin\hat{a}$, $\sin\hat{b}$, $\sin\hat{c}$. Запишем теорему косинусов для этого треугольника в виде

$$\cos(\pi-\hat{a})=\frac{\sin^2(\pi-\hat{b})+\sin^2(\pi-\hat{c})-\sin^2(\pi-\hat{a})}{2\sin(\pi-\hat{b})\sin(\pi-\hat{c})}.$$

Подставляя в это равенство выражения через параметры, получаем цепочку следствий

$$\sqrt{1-\frac{a}{p}}=\frac{b/p+c/p-a/p}{2\tilde{b}\tilde{c}/p} \quad\Longrightarrow\quad \frac{p-a}{p}=\frac{(p-a)^2}{bc} \quad\Longrightarrow$$

$$(p-a)p=bc \quad\Longrightarrow\quad a^2=(b-c)^2 \quad\Longrightarrow\quad (p-b)(p-c)=0 \quad\Longrightarrow\quad D=0.$$

Итак, мы показали, что чётное число переменных входит в уравнения (3) с разными знаками. Если таких переменных две, то можно считать без ограничения общности, что знаки + выбраны у них в одном и

том же уравнении. Поэтому имеем 4 системы линейных уравнений, дающие все решения исходной системы (1). Запишем две из них:

$$\begin{cases} \hat u+ \quad \hat v+ \quad \hat c= \quad \pi \pmod{2\pi},\\ \hat v+ \quad \hat w+ \quad \hat a= \quad \pi \pmod{2\pi},\\ \hat w+ \quad \hat u+ \quad \hat b= \quad \pi \pmod{2\pi}, \end{cases} \qquad \begin{cases} \hat u+ \quad \hat v- \quad \hat c= \quad \pi \pmod{2\pi},\\ \hat v+ \quad \hat w+ \quad \hat a= \quad \pi \pmod{2\pi},\\ \hat w+ \quad \hat u+ \quad \hat b= \quad \pi \pmod{2\pi}, \end{cases} \tag{4}$$

остальные получаются из правой системы в (4) циклическим сдвигом.

Решая системы (4), имеем

$$\begin{aligned} \hat u &= \frac{-\sigma_a+\sigma_b+\sigma_c}{2}\pi-(\hat p-\hat a), & \hat u &= \frac{-\sigma_a+\sigma_b+\sigma_c}{2}\pi+ \quad (\hat p-\hat b),\\ \hat v &= \frac{\sigma_a-\sigma_b+\sigma_c}{2}\pi-(\hat p-\hat b), & \hat v &= \frac{\sigma_a-\sigma_b+\sigma_c}{2}\pi+ \quad (\hat p-\hat a),\\ \hat w &= \frac{\sigma_a+\sigma_b-\sigma_c}{2}\pi-(\hat p-\hat c), & \hat w &= \frac{\sigma_a+\sigma_b-\sigma_c}{2}\pi- \qquad \hat p, \end{aligned}$$

где использовано обозначение $\hat p = \frac12(\hat a+\hat b+\hat c)$, а переменные $\sigma_a, \sigma_b, \sigma_c$ принимают значения ±1. Легко видеть, что с точностью до общего для всех переменных кратного π эти выражения равны

$$\begin{aligned} \hat u &= \frac{\pi}{2}- \quad (\hat p-\hat a), & \hat u &= \frac{\pi}{2}- \quad (\hat p-\hat a),\\ \hat v &= \frac{\pi}{2}- \quad (\hat p-\hat b), & \hat v &= \frac{\pi}{2}- \quad (\hat p-\hat b),\\ \hat w &= \frac{\pi}{2}- \quad (\hat p-\hat c), & \hat w &= \frac{\pi}{2}- \qquad \hat p. \end{aligned}$$

Сдвиг всех угловых переменных на π приводит к изменению знака у переменных u, v, w. Поскольку $u = \sqrt{p}\sin\hat u, \dots$, то получаем окончательные выражения для всех решений системы (1) через тригонометрические функции угловых параметров (выражение перед матрицей применяется к каждому элементу матрицы, каждая строчка матрицы даёт два решения системы, отличающиеся изменением знака у всех неизвестных):

$$(u,v,w) = \pm\sqrt{p}\cos\begin{pmatrix} \hat p-\hat a & \hat p-\hat b & \hat p-\hat c\\ \hat p & \hat p-\hat c & \hat p-\hat b\\ \hat p-\hat c & \hat p & \hat p-\hat a\\ \hat p-\hat b & \hat p-\hat a & \hat p \end{pmatrix}. \tag{5}$$

Обратим внимание на то, что хотя $\sin\hat p$ и $\cos\hat p$ определены лишь с точностью до знака (половинный угол известного угла), решений это не добавляет.

Заметим, что формулы (5) могут быть записаны как алгебраические функции от исходных параметров.

4. Формулы для радиусов окружностей в задаче Мальфатти

Теперь покажем, как из решений основной системы получаются выражения для радиусов окружностей в задаче Мальфатти.

Будем использовать обозначения, введённые на рис. 2.

Для любого решения задачи Мальфатти выполняются равенства

$$\begin{cases} a = BC = \pm\, BB_C \pm B_CC_B \pm C_BC, \\ b = AC = \pm\, AA_C \pm A_CC_A \pm C_AC, \\ c = AB = \pm\, AA_B \pm A_BB_A \pm B_AB. \end{cases} \tag{6}$$

Выбор знаков зависит от рассматриваемой конфигурации.

Простые вычисления показывают, что $B_CC_B = 2\sqrt{r_br_c}$ (и ещё два аналогичных равенства получаются циклическим сдвигом $A \to B \to C \to A$). Коэффициент пропорциональности между r_b и BB_A зависит от величины угла $\angle ABC = \beta$ (он равен $\operatorname{tg}\frac{\beta}{2}$, если окружность вписана во внутренний угол или вертикальный с ним, в противном случае он равен $\operatorname{ctg}\frac{\beta}{2}$). Введём неизвестные u, v, w, модули которых

$$|u| = \sqrt{AA_B} = \sqrt{AA_C};\ |v| = \sqrt{BB_C} = \sqrt{BB_A};\ |w| = \sqrt{CC_A} = \sqrt{CC_B},$$

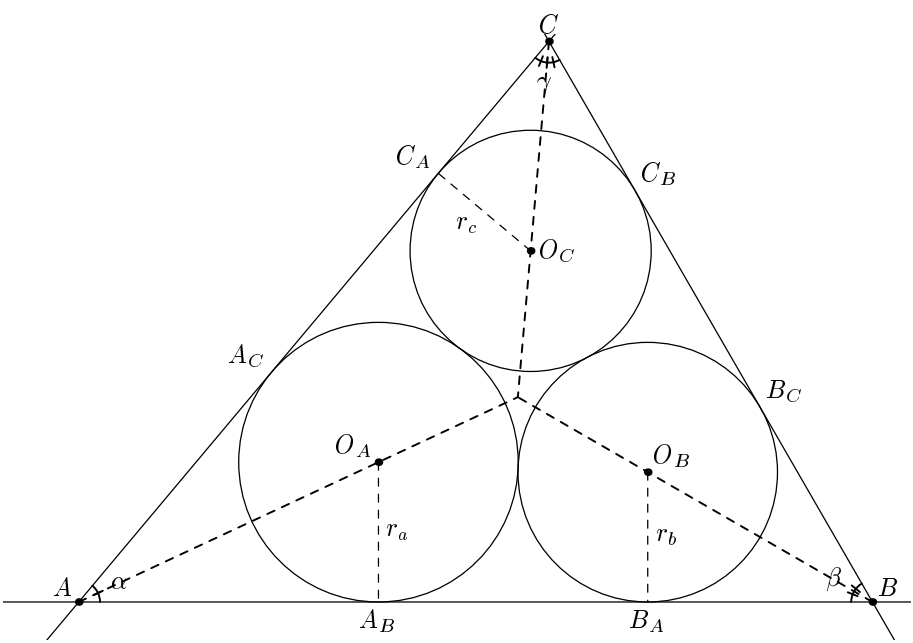


Рис. 2. *Случай, когда в* (6) *все знаки* $+$.

а аргументы будут выбираться для разных конфигураций по-разному из возможных значений $k\frac{\pi}{4}$, $k = 0 \dots 3$. Тогда

$$A_B B_A = 2\sqrt{r_a r_b} = 2|u||v|\sqrt{\operatorname{tg}^{\pm 1}\frac{\alpha}{2}\operatorname{tg}^{\pm 1}\frac{\beta}{2}}$$
$$B_C C_B = 2\sqrt{r_b r_c} = 2|v||w|\sqrt{\operatorname{tg}^{\pm 1}\frac{\beta}{2}\operatorname{tg}^{\pm 1}\frac{\gamma}{2}}$$
$$C_A A_C = 2\sqrt{r_c r_a} = 2|w||u|\sqrt{\operatorname{tg}^{\pm 1}\frac{\gamma}{2}\operatorname{tg}^{\pm 1}\frac{\alpha}{2}}.$$

Тангенс или котангенс в этих формулах выбираются в зависимости от того, вписана ли соответствующая окружность во внутренний (или вертикальный ему) угол, либо же в смежный с ним.

Выразим тангенсы половинных углов треугольника через его стороны. Для этого воспользуемся формулой $S^2 = p^2 r^2 = p(p-a)(p-b)(p-c)$. Получаем

$$\operatorname{tg}\frac{\alpha}{2} = \sqrt{\frac{(p-b)(p-c)}{p(p-a)}}$$

и аналогичные формулы для двух других углов. Из этих формул получаются следующие соотношения (опять-таки приводится одно соотношение, из которого перестановками сторон можно получать аналогичные):

$$\sqrt{\operatorname{tg}\frac{\alpha}{2}\operatorname{tg}\frac{\beta}{2}} = \sqrt{\frac{p-c}{p}} = \sqrt{1-\frac{c}{p}}, \quad \sqrt{\operatorname{ctg}\frac{\alpha}{2}\operatorname{tg}\frac{\beta}{2}} = \sqrt{\frac{p-a}{p-b}}, \tag{7}$$

$$\sqrt{\operatorname{ctg}\frac{\alpha}{2}\operatorname{ctg}\frac{\beta}{2}} = \sqrt{\frac{p}{p-c}}. \tag{8}$$

Подставляя полученные выражения в систему (6), будем получать системы уравнений относительно $|u|, |v|, |w|$ и сводить их к системе (1) выбором аргументов неизвестных и преобразованиями параметров.

4.1. Первая восьмёрка решений

Пусть все три окружности вписаны во внутренние углы. Тогда коэффициенты при средних слагаемых в левых частях системы (6) имеют вид (7), поэтому система (6) сразу имеет вид (1) с точностью до знаков слагаемых в левых частях уравнений.

Положительные знаки в системе (6) соответствуют первой конфигурации на рис. 1 (см. стр. 143). Для этой конфигурации выполняется система (1), если считать неизвестные u, v, w положительными. Уравнения (6)

для второй конфигурации можно привести к виду (1), поменяв знак у одной из неизвестных (это изменит два знака в средних слагаемых).

Решим систему (1), как описано выше. Одновременная замена знаков не меняет геометрического решения (мы используем квадраты неизвестных и попарные произведения). Поэтому получается четыре различных решения. Но как было показано выше, решения, соответствующие первой и второй конфигурациям, удовлетворяют системе (1) при различных значениях переменных. Таким образом, формула (5) даёт решения для этих конфигураций.

Чтобы найти решения, соответствующие третьей и четвёртой конфигурациям на рис. 1, заметим, что этим конфигурациям отвечает выбор отрицательных знаков у средних слагаемых в (6) и, как и в предыдущем случае, перемена знака у одной из неизвестных. Изменение знака у средних слагаемых можно выразить в терминах угловых параметров как сдвиг на π значений всех угловых параметров. Поэтому (выбираем подходящий знак!) в (5) синусы и косинусы поменяются местами.

4.2. Остальные решения

Нетрудно убедиться, разглядывая рис. 1, что во всех остальных конфигурациях одна окружность вписана во внутренний (или вертикальный ему) угол, а две другие — во внешние углы треугольника, образованного прямыми. Понять это можно и с помощью рассуждения — если одна окружность вписана во внешний угол, а две другие — нет, то одну из пар окружностей будет разделять одна из прямых l_1, l_2, l_3. Аналогично в случае, когда все три окружности вписаны во внешние углы.

Итак, все остальные решения разбиваются на три группы, в зависимости от того, в какой из внутренних углов вписана окружность. Мы будем считать без ограничения общности, что вершина, соответствующая этому углу, — это B. Две другие серии решений получаются из данной циклическими перестановками вершин.

Перепишем систему (6) с учётом соотношений (7–8). Получаем

$$\begin{cases} c = \pm|u|^2 \pm 2|u||v|\sqrt{\operatorname{tg}\dfrac{\beta}{2}\operatorname{ctg}\dfrac{\alpha}{2}} \pm |v|^2 = \pm|u|^2 \pm 2|u||v|\sqrt{\dfrac{p-a}{p-b}} \pm |v|^2, \\ a = \pm|v|^2 \pm 2|u||v|\sqrt{\operatorname{ctg}\dfrac{\gamma}{2}\operatorname{tg}\dfrac{\beta}{2}} \pm |w|^2 = \pm|v|^2 \pm 2|u||v|\sqrt{\dfrac{p-c}{p-b}} \pm |w|^2, \\ b = \pm|w|^2 \pm 2|w||u|\sqrt{\operatorname{ctg}\dfrac{\gamma}{2}\operatorname{ctg}\dfrac{\alpha}{2}} \pm |u|^2 = \pm|w|^2 \pm 2|w||u|\sqrt{\dfrac{p}{p-b}} \pm |u|^2. \end{cases} \tag{9}$$

Рассмотрим конфигурацию 5 на рис. 1. Расставим для неё знаки в (9) и введем новые параметры $a' = -a,\ b' = b,\ c' = -c$. Получим

$$\begin{cases} c' = |u|^2 + 2|u||v|\sqrt{-\left(1 - \dfrac{c'}{p'}\right)} - |v|^2, \\ a' = -|v|^2 + 2|u||v|\sqrt{-\left(1 - \dfrac{a'}{p'}\right)} + |w|^2, \\ b' = |w|^2 + 2|w||u|\sqrt{1 - \dfrac{b'}{p'}} + |u|^2. \end{cases} \tag{10}$$

Если считать, что неизвестные u, w — положительные, а v — чисто мнимая с отрицательной мнимой частью, то система (10) превращается в систему (1) относительно параметров a', b', c' (эту замену можно проинтерпретировать как переход к комплексному треугольнику со сторонами $(-a, b, -c)$).

Решения этой системы задаются формулами (5). Нам нужно убедиться, что среди них есть решение для конфигурации 5. Для этого выразим решения (5) через гиперболические функции[3)] от мнимых частей угловых параметров.

Пусть

$$\hat{a} = \varphi_a + it_a,\ \hat{b} = \varphi_b + it_b,\ \hat{c} = \varphi_c + it_c.$$

Тогда, используя формулу Эйлера, для углового параметра $\hat{c}$ можно записать следующие равенства

$$\begin{aligned} \cos\hat{c} &= -i\sqrt{\frac{p-a}{p-b}} &&= \operatorname{ch} t_c \cos\varphi_c - i \operatorname{sh} t_c \sin\varphi_c, \\ \sin\hat{c} &= \sqrt{\frac{c}{p-b}} &&= \operatorname{ch} t_c \sin\varphi_c + i \operatorname{sh} t_c \cos\varphi_c. \end{aligned}$$

Из этих и аналогичных равенств получаем выражения для вещественной и мнимой части $\hat{c}$ и $\hat{a}$:

$$\begin{aligned} \varphi_c &= \frac{\pi}{2},\ t_c = \operatorname{arcch}\sqrt{\frac{c}{p-b}} = \operatorname{arcsh}\sqrt{\frac{p-a}{p-b}}, \\ \varphi_a &= \frac{\pi}{2},\ t_a = \operatorname{arcch}\sqrt{\frac{a}{p-b}} = \operatorname{arcsh}\sqrt{\frac{p-c}{p-b}}. \end{aligned}$$

[3)] Напомним, что гиперболические функции определяются так: $\operatorname{ch} x \stackrel{\text{def}}{=} (e^x + e^{-x})/2$, $\operatorname{sh} x \stackrel{\text{def}}{=} (e^x - e^{-x})/2$. Подробнее о свойствах гиперболических функций можно прочитать в [8].

Для углового параметра $\hat{b}$ формулы несколько отличаются:

$$\cos\hat{b} = -\sqrt{\frac{p}{p-b}} = \operatorname{ch} t_b \cos\varphi_b - i \operatorname{sh} t_b \sin\varphi_b,$$
$$\sin\hat{b} = -i\sqrt{\frac{b}{p-b}} = \operatorname{ch} t_b \sin\varphi_b + i \operatorname{sh} t_b \cos\varphi_b,$$
$$\varphi_b = \pi,$$
$$t_b = -\operatorname{arcch}\sqrt{\frac{p}{p-b}} = -\operatorname{arcsh}\sqrt{\frac{b}{p-b}}.$$

Через t обозначим $(t_a + t_b + t_c)/2$. Получаем искомое выражение формул (5) через мнимые части угловых параметров.

$$(u,v,w) = \pm\sqrt{p-b}\begin{pmatrix} \operatorname{sh}(t-t_a) & i\operatorname{ch}(t-t_b) & \operatorname{sh}(t-t_c) \\ -i\operatorname{ch} t & \operatorname{sh}(t-t_c) & i\operatorname{ch}(t-t_b) \\ \operatorname{sh}(t-t_c) & -i\operatorname{ch} t & \operatorname{sh}(t-t_a) \\ i\operatorname{ch}(t-t_b) & \operatorname{sh}(t-t_a) & -i\operatorname{ch} t \end{pmatrix}. \tag{11}$$

Первая строка соответствует решению для конфигурации 5. Чтобы в этом убедиться, достаточно проверить, что $(t-t_a)$, $(t-t_c)$ отрицательны. Элементарными вычислениями с учётом неравенства треугольника для (a,b,c) проверяется, что $\operatorname{sh}(t_b+t_c) < 0 < \operatorname{sh} t_a$, откуда в силу монотонности гиперболического синуса следует, что $t_b + t_c < t_a$. Второе неравенство доказывается аналогично.

Сравнивая аргументы неизвестных в (11) и требуемые наборы знаков в системах (6) для конфигураций 6 и 7, убеждаемся, что остальные три строки также дают решения задачи Мальфатти: третья — для конфигурации 6, вторая — для конфигурации 7, когда отмеченная на рис. 1 вершина $H = C$, четвёртая — для конфигурации 7, когда $H = A$.

Как и в случае первой восьмёрки решений, выписывая системы (6) для конфигураций 8–10, можно проверить, что при подходящем выборе аргументов неизвестных u, v, w эти системы совпадают с системой (10), если изменить знаки коэффициентов при средних слагаемых. Это означает сдвиг угловых параметров на π (напомним, что знаки синусов угловых

параметров выбираются произвольно). Получаем следующие решения:

$$(u,v,w) = \pm\sqrt{p-b}\begin{pmatrix} i\operatorname{ch}(t-t_a) & -\operatorname{sh}(t-t_b) & i\operatorname{ch}(t-t_c) \\ -\operatorname{sh}t & i\operatorname{ch}(t-t_c) & -\operatorname{sh}(t-t_b) \\ i\operatorname{ch}(t-t_c) & -\operatorname{sh}t & i\operatorname{ch}(t-t_a) \\ -\operatorname{sh}(t-t_b) & i\operatorname{ch}(t-t_a) & -\operatorname{sh}t \end{pmatrix}. \quad (12)$$

В этих формулах первая строка соответствует конфигурации 8, а третья — конфигурации 9. В этом можно убедиться, сравнивая значения $|v|$ в одном и другом случае. Вторая и четвёртая строки соответствуют двум случаям конфигурации 10.

5. Геометрическая интерпретация решений

В заключение коротко опишем геометрическую интерпретацию решений.

Для первой восьмёрки решений рассмотрим сферу радиуса $\sqrt{p}/2$ и сферический треугольник $\tilde{\Delta}$ на ней, стороны которого стягиваются хордами длин $\sqrt{a}$, $\sqrt{b}$, $\sqrt{c}$. Плоскости, проходящие через центр сферы и стороны $\tilde{\Delta}$ разбивают сферу на 8 сферических треугольников. Каждый из этих треугольников определяет одно из решений. А именно, впишем в треугольник окружность, тогда расстояния между вершинами треугольника и точками касания дают модули искомых неизвестных u, v, w.

В остальных случаях нужно провести аналогичное построение в пространстве, снабжённом псевдометрикой $d^2 = x^2 - y^2 - z^2$. Роль сфер в этом пространстве играют двуполосные гиперболоиды (псевдосферы), а их центральные плоские сечения суть «большие круги». Три таких сечения разбивают псевдосферу на 8 областей, одна из них имеет линейные размеры $\sqrt{a'}$, $\sqrt{b'}$, $\sqrt{c'}$. В псевдометрике они являются либо вещественными, либо чисто мнимыми числами. Окружность, вписанная в псевдосферический треугольник, определяется точками касания образующих его дуг с общей касательной плоскостью. Модули расстояний от точек касания вписанной окружности до вершин треугольника дают модули искомых неизвестных.

Список литературы

[1] *Malfatti.* Memoirie di matematica. Tomo X. Parte I. Modena. 1803.

[2] *Steiner J.* Einige geometrische Betrachtungen. // Crelle J. Parte I. 1826.

[3] *Petersen J.* Methodes et theories pour la resolution des problemes des constructions geometriques. Paris: Gautier-Villars. 1880.

[4] *Адлер А.* Теория геометрических построений. Одесса: изд-во "Mathesis". 1910.

[5] *Шклярский Д. О., Ченцов Н. Н., Яглом И. М.* Избранные задачи и теоремы математики, ч. 2. Геометрия (планиметрия). М.: Гостехиздат. 1952. (Серия «Библиотека математического кружка».)

[6] *Беленький В. З., Заславский А. А.* О задаче Мальфатти. // Квант. 1994. №4.

[7] *Адамар Ж.* Элементарная геометрия, часть I. Планиметрия. М.: Учпедгиз. 1948.

[8] *Шерватов В. Г.* Гиперболические функции. М.: Гостехиздат. 1954. (Серия «Популярные лекции по математике».)

Наш семинар: математические сюжеты

Всякое ли чебышёвское множество выпукло?

А. Р. Алимов*

Пусть M — непустое замкнутое подмножество плоскости и x — точка, не принадлежащая ему. Тогда для точки x всегда существует точка из M, ближайшая к ней (это доказывается ниже), т. е. такая точка $y_0 \in M$, что расстояние от x до y_0 не больше, чем расстояние от x до любой точки $y \in M$. (Таких ближайших к x точек может быть, вообще говоря, много.)

Если M — замкнутый круг и точка x лежит вне M, то ближайшая к x точка y_0 единственна, она лежит на пересечении окружности, ограничивающей круг M, и луча, начинающегося в центре круга и проходящего через точку x. Но если M состоит всего из двух точек y_1 и y_2, то возможны два случая:

– точка x не лежит на серединном перпендикуляре l к отрезку $[y_1; y_2]$, ближайшая к x единственна — или y_1, или y_2;

– точка x лежит на l, ближайшими к x будут *обе* точки y_1 и y_2.

Точки, для которых ближайшая точка неединственна, напоминают об известном *буридановом осле*, который стоял между двух равноудалённых от него мешков с овсом и погиб от голода, так и не решив, какой из мешков находится к нему ближе.

Возникает вопрос: *как описать все замкнутые множества M, для которых ближайшая из M к любой точке x единственна?*

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект № 96-01-00212) и программы «Ведущие научные школы» (проект № 96-15-96102).

Вначале мы рассмотрим этот вопрос в евклидовой плоскости $\mathbb{R}^2$, а затем — в более общем n-мерном евклидовом пространстве $\mathbb{R}^n$.

Для точек x, y с координатами $x = (\alpha_1, \alpha_2)$ и $y = (\beta_1, \beta_2)$ на плоскости $\mathbb{R}^2$ определим число $\langle x, y\rangle = \alpha_1\beta_1 + \alpha_2\beta_2$, называемое *скалярным произведением* векторов x и y. Из определения $\langle\cdot,\cdot\rangle$ сразу следует, что $\langle x, y\rangle = \langle y, x\rangle$ и $\langle \alpha x + \beta y, z\rangle = \alpha\langle x, z\rangle + \beta\langle y, z\rangle$ при $\alpha, \beta \in \mathbb{R}$. Под *расстоянием* между ними мы будем понимать обычное евклидово расстояние

$$\|x - y\| = \sqrt{\langle x - y, x - y\rangle} = \sqrt{(\alpha_1 - \beta_1)^2 + (\alpha_2 - \beta_2)^2}.$$

Обобщением расстояния между точками является понятие *наилучшего приближения* или *расстояния* $\rho(x, M)$ *от заданной точки* x *до заданного множества* M

$$\rho(x, M) = \inf_{y \in M} \|x - y\|.$$

Под *элементом наилучшего приближения* или *ближайшей точкой* для заданной точки x мы будем понимать ту точку $y_0 \in M$, для которой $\|x - y_0\| = \rho(x, M)$, т. е. $\|x - y_0\| \leqslant \|x - y\|$ для любых $y \in M$. Множество всех ближайших точек из M для заданной точки x обозначается Px.

Определение. Непустое множество M называется *чебышёвским*, если любая точка x имеет точно одну ближайшую в M:

$$\forall x \quad Px \text{ состоит из одной точки.}$$

Если M — чебышёвское множество, то отображение P, сопоставляющее точке x её ближайшую точку Px из M, называется *метрической проекцией* на множество M.

Простыми примерами чебышёвских множеств служат замкнутый круг, отрезок или прямая на плоскости $\mathbb{R}^2$.

Используя введённое понятие чебышёвского множества, переформулируем поставленный выше вопрос: *описать все чебышёвские множества в* $\mathbb{R}^2$ (в $\mathbb{R}^n$).

Первые значительные результаты, относящиеся к чебышёвским множествам, были получены в теории приближения функций одним из основателем этой теории П. Л. Чебышёвым. Название «чебышёвское множество» было дано позднее С. Б. Стечкиным в честь Чебышёва.

Первыми, кто исследовал геометрические свойства чебышёвских множеств, были Л. Бунт и Т. Моцкин. В середине 30-х годов Бунт ответил на поставленный вопрос для случая $\mathbb{R}^n$ (см. обзоры [2] и [3]) и нашёл геометрическое свойство, характеризующее чебышёвские множества. Таким

геометрическим свойством оказалась выпуклость. Множество M называется *выпуклым*, если для любых точек $x, y \in M$ множество M содержит отрезок $[x; y]$, их соединяющий. Например, круг на плоскости — выпуклое множество, а окружность — нет.

Теорема (Л. Бунт). *Множество в евклидовой плоскости (в пространстве $\mathbb{R}^n$) является чебышёвским тогда и только тогда, когда оно замкнуто и выпукло.*

Замкнутость чебышёвского множества очевидна. Если бы оно было незамкнутым, то нашлась бы предельная точка множества (т. е. точка, расположенная на нулевом расстоянии от множества), ему не принадлежащая, но это противоречило бы чебышёвости множества. Поэтому далее все рассматриваемые множества будут предполагаться замкнутыми.

Достаточность в теореме почти очевидна и была известна очень давно. Существенно более трудно доказать *необходимость* выпуклости. Мы приводим несколько доказательств этого замечательного утверждения. Доказательство достаточности и первые два доказательства необходимости для наглядности будут проведены для множеств в евклидовой плоскости $\mathbb{R}^2$.

Введём несколько обозначений. Если $x \in \mathbb{R}^2$ и $r > 0$, то положим
$B(x, r) = \{y \in \mathbb{R}^2 \mid \|x - y\| \leqslant r\}$ — круг с центром x и радиусом r,
$\mathring{B}(x, r) = \{y \in \mathbb{R}^2 \mid \|x - y\| < r\}$ — открытый круг с центром x и радиусом r, («внутренность» круга $B(x, r)$),
$S(x, r) = \{y \in \mathbb{R}^2 \mid \|x - y\| = r\}$ — окружность с центром x и радиусом r.

Доказательство достаточности.
Пусть множество M, расположенное в евклидовой плоскости, выпукло и замкнуто и точка x не лежит в M. Докажем, что x имеет в точности одну ближайшую точку из M.

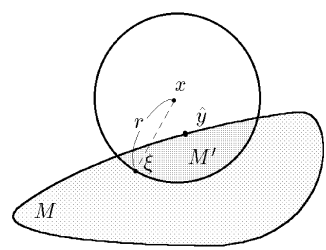


Рис. 1.

Доказательство существования ближайшего элемента извлечём из теоремы Вейерштрасса: *непрерывная функция на компакте*[1] *достигает своего минимума.* Возьмём любую точку $\xi \in M$ и положим $r = \|x - \xi\|$. Тогда (см. рис. 1) множество $M' = M \cap B(x, r)$ выпукло, замкнуто (как пересечение двух выпуклых и замкнутых множеств) и ограничено (ибо лежит в круге радиуса r). Следовательно, оно компактно.

[1] В $\mathbb{R}^2$ и в $\mathbb{R}^n$ компакт — это в точности замкнутое ограниченное множество.

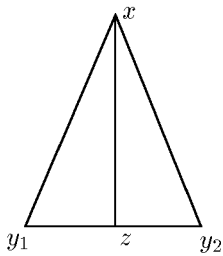


Рис. 2.

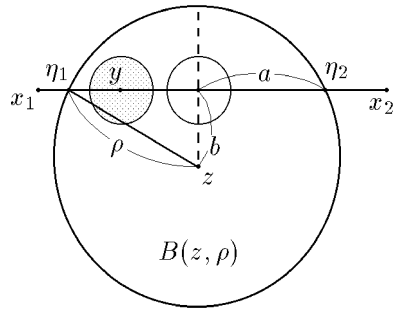


Рис. 3.

Функция $f(y) = \|x - y\| = \sqrt{(\alpha_1 - \beta_1)^2 + (\alpha_2 - \beta_2)^2}$ непрерывна. Значит, она достигает минимума на M' в некоторой точке $\hat{y}$, т. е. $\|x - \hat{y}\| \leqslant \leqslant \|x - \eta\|$ для всех $\eta \in M'$. А вне M' любая точка отстоит от x на расстояние, большее $\|x - \xi\|$. Значит, $\hat{y}$ — ближайшая точка к x в M.

Доказательство единственности. Пусть M — выпуклое замкнутое множество. Допустим, что для некоторой точки $x \notin M$ найдутся две ближайшие точки $y_1, y_2 \in M$. Поскольку M выпукло, то точка $z = (y_1 + + y_2)/2$ — середина отрезка $[y_1; y_2]$ — принадлежит M.

В (невырожденном) равнобедренном треугольнике с вершинами x, y_1, y_2 (см. рис. 2) точка z является основанием высоты, опущенной из x на сторону $[y_1; y_2]$, т. е. её длина *меньше* длины боковой стороны — числа $\|x - y_1\| = \|x - y_2\|$. Итак, мы нашли точку $z \in M$, более близкую к x, чем y_1 и y_2, что противоречит выбору точек y_1 и y_2. Следовательно, исходное предположение о наличии у x двух ближайших точек было неверно, поэтому M — чебышёвское множество. □

1. Доказательство Л. Бунта

(См. [4].) Допустим, что чебышёвское множество M невыпукло. Тогда найдутся точки $x_1, x_2 \in M$ и точка y на отрезке $[x_1; x_2]$ такие, что y не лежит в M. Поскольку M замкнуто и $y \notin M$, то найдётся число $\varepsilon > 0$ такое, что круг $B(y, \varepsilon)$ не пересекается с M.

Рассмотрим совокупность Ω *всех* замкнутых кругов $B(z, \rho)$, каждый из которых содержит круг $B(y, \varepsilon)$ и не пересекается внутренностью с M ($B(y, \varepsilon) \subset B(z, \rho)$ и $\mathring{B}(z, \rho) \cap M = \varnothing$). Ясно, что Ω замкнуто. Покажем, что радиусы ρ этих кругов ограничены.

Действительно, пусть $B(z, \rho) \in \Omega$ пересекает отрезок $[x_1; x_2]$ в двух точках η_1 и η_2, расстояние между которыми $2a$ (см. рис. 3). Обозначим

через b расстояние от z до хорды $[\eta_1;\eta_2]$. Ясно, что $b+\varepsilon\leqslant\rho=\sqrt{a^2+b^2}$, откуда $b^2+2b\varepsilon+\varepsilon^2\leqslant a^2+b^2$ и $b\leqslant(a^2-\varepsilon^2)/(2\varepsilon)$, значит, $\rho^2=a^2+b^2\leqslant\leqslant a^2+\big((a^2-\varepsilon^2)/(2\varepsilon)\big)^2$, что и требовалось.

Если $B(z,\rho)\in\Omega$, то для его центра z выполнены два условия:
– проекция на отрезок $[x_1;x_2]$ лежит между x_1 и x_2;
– расстояние от z до прямой, соединяющей x_1 и x_2, не больше, чем $(a^2-\varepsilon^2)/(2\varepsilon)$.

Получилось, что совокупность (z,ρ), соответствующих $B(z,\rho)\in\Omega$, есть компакт. Значит, по теореме Вейерштрасса в Ω найдётся круг $B(z_0,\rho_0)$ максимального радиуса. При этом обязательно

$$B(z_0,\rho_0)\cap M\neq\varnothing,$$

иначе в Ω найдётся круг $B(z_0,\rho_1)$ *большего, чем* ρ_0, радиуса (достаточно положить $\rho_1=\min_{v\in M}\|v-z_0\|>\rho_0$, тогда $B(z_0,\rho_1)\cap M=\varnothing$ и $B(y,\varepsilon)\subset\subset B(z_0,\rho_0)\subset B(z_0,\rho_1)$). Пусть теперь $y_0\in M$ — точка, ближайшая к z_0, тогда $\rho(z_0,M)=\|z_0-y_0\|=\rho_0$.

Поскольку $B(y,\varepsilon)\subset B(z_0,\rho_0)$, то граничная окружность $S(z_0,\rho_0)$ круга $B(z_0,\rho_0)$ или не имеет с $B(y,\varepsilon)$ ни одной общей точки, или имеет одну общую точку w (рис. 4), которая, так как $B(y,\varepsilon)\cap M=\varnothing$, отлична от точки $\{y_0\}=M\cap S(z_0,\rho_0)$. Если мы теперь сдвинем круг $B(z_0,\rho_0)$ на *малое* расстояние в направлении вектора $\overrightarrow{y_0z_0}$ (в случае $S(z_0,\rho_0)\cap B(y,\varepsilon)=\varnothing$) или в направлении вектора $\overrightarrow{y_0w}$ (в случае $S(z_0,\rho_0)\cap B(y,\varepsilon)=\{w\}$), то касание сдвинутого круга $B(z_0,\rho_0)$ и множества M исчезнет.

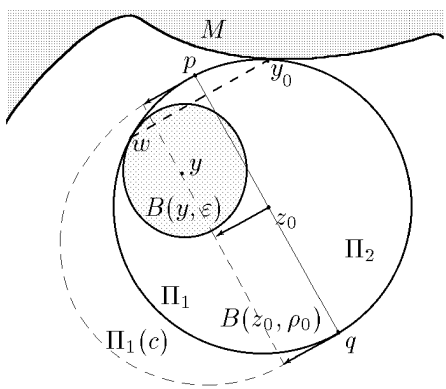


Рис. 4.

Действительно, рассмотрим второй случай (рис. 4). Пусть точки $p,q\in\in S(z_0,\rho_0)$ выбраны так, что диаметр $[p;q]$ перпендикулярен хорде $[w;y_0]$. Диаметр $[p;q]$ разбивает круг на два полукруга: Π_1 и Π_2, при этом считаем, что $w\in\Pi_1$, $y_0\in\Pi_2$. Поскольку M замкнуто и не пересекается с Π_1, то расстояние от любой точки полукруга Π_1 до множества M больше некоторого числа $a>0$. Поэтому найдётся число $b>0$, что для любого c, $0\leqslant c\leqslant b$, сдвиг $\Pi_1(c)=\Pi_1+c\cdot\overrightarrow{y_0w}$ полукруга Π_1 на вектор $c\cdot\overrightarrow{y_0w}$ не пересекается с M. Далее, при сдвиге на вектор $b\cdot\overrightarrow{y_0w}$ любая точка правого полукруга Π_2 перейдёт в некоторую точку полукруга $\Pi_1(c)$

для некоторого $c \in [0; b]$ и, по доказанному выше, не будет лежать в M. Итак, сдвинутый круг $B(z_0', \rho_0) = B(z_0, \rho_0) + c \cdot \overrightarrow{y_0 w}$ не пересекается с M ($z_0' = z_0 + c \cdot \overrightarrow{y_0 w}$), т. е.

$$B(y, \varepsilon) \subset B(z_0', \rho_0), \quad B(z_0', \rho_0) \cap M = \varnothing.$$

Значит, $B(z_0', \rho_0) \in \Omega$. Но тогда $B(z_0', \rho_2) \in \Omega$ для $\rho_2 = \min_{v \in M} \|v - z_0'\|$, что ввиду $\rho_2 > \rho_0$ противоречит максимальности ρ_0. Это показывает, что допущение невыпуклости M ведёт к противоречию.

Первый случай рассматривается аналогично. □

2. Доказательство В. Кли – В. И. Бердышева – Л.П. Власова

(См. [3, 8].) Ключевую роль в этом доказательстве играет теорема Брауэра[2)] о существовании неподвижной точки при непрерывном отображении круга в себя (см. [5], [6]): *при непрерывном отображении f круга $B(x, r)$ в себя существует неподвижная точка; т. е. найдётся точка $z_0 \in B(x, r)$ такая, что $f(z_0) = z_0$.* Эта важная теорема также верна и при непрерывном отображении произвольного выпуклого компакта из $\mathbb{R}^n$ в себя.

Условие выпуклости в теореме Брауэра важно. Например, любой поворот окружности на угол, отличный от $360n^\circ$, $n \in \mathbb{N}$, *не* имеет неподвижной точки.

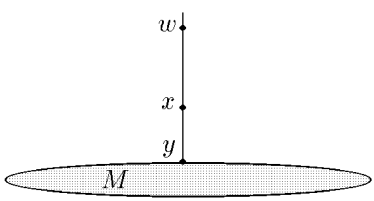


Рис. 5.

Определение. Чебышёвское множество M называется *чебышёвским солнцем*, если (см. рис. 5) для любой точки x, не лежащей в M, любая точка w из луча, начинающегося в $y = Px$ и проходящего через x, имеет точку y своей ближайшей из M.

Мы докажем теорему Бунта, показав, что всякое чебышёвское множество является чебышёвским солнцем и что всякое чебышёвское солнце выпукло.

Лемма 1. *Чебышёвское множество на плоскости* (в $\mathbb{R}^n$) *является чебышёвским солнцем.*

[2)] Бердышев использует более слабое, чем теорема Брауэра, утверждение. Его доказательство в 1960 г. получило золотую медаль на конкурсе студенческих работ.

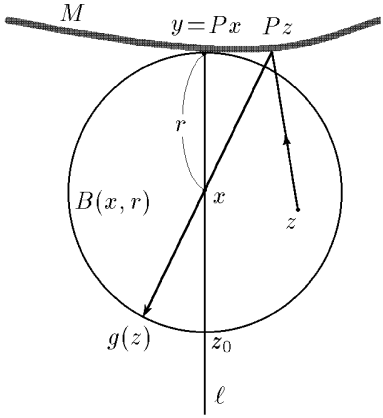


Рис. 6.

Доказательство леммы 1. Пусть M — чебышёвское множество и $x \notin M$. Посредством ℓ обозначим луч, начинающийся в $y = Px$ и проходящий через x. Положим $r = \|x - y\|$ и обозначим $B = B(x, \|x - y\|)$. Определим отображение $g : B \to B$ по формуле

$$g(z) = x + \frac{\|x - y\|}{\|x - Pz\|}(x - Pz),\ z \in B$$

(точка $g(z)$ лежит на пересечении луча, исходящего из x в направлении $\overrightarrow{Pz\,x}$, с окружностью $S(x, r)$) (см. рис. 6)).

Докажем, что отображение g непрерывно. Для этого покажем, что метрическая проекция P на чебышёвское множество непрерывна.

Сначала заметим, что если $u, v \in \mathbb{R}^2$ $(u, v \in \mathbb{R}^n)$, то

$$\rho(u, M) \leqslant \|u - v\| + \rho(v, M), \quad |\rho(u, M) - \rho(v, M)| \leqslant \|u - v\|. \qquad (1)$$

Действительно, для любого $s \in M$ в силу неравенства треугольника имеем $\|u - s\| \leqslant \|u - v\| + \|v - s\|$. Возьмём нижнюю грань по $s \in M$ сначала в левой части неравенства, а затем — в правой. Получим $\rho(u, M) \leqslant \leqslant \|u - v\| + \rho(v, M)$. Поменяв местами u и v в предыдущем рассуждении, получим $\rho(v, M) \leqslant \|u - v\| + \rho(u, M)$. Поэтому

$$|\rho(v, M) - \rho(u, M)| \leqslant \|u - v\|.$$

Теперь предположим, что метрическая проекция P разрывна в некоторой точке u, т. е. найдутся число $\varepsilon > 0$ и последовательность $\{u_n\}_{n\in\mathbb{N}}$, $u_n \to u$, такие, что $\|v_n - v\| \geqslant \varepsilon$ для всех $n \in \mathbb{N}$, где $Pu_n = v_n$, $Pu = v$ (можем считать, что $u \neq v$). В силу (1) последовательность $\{v_n\}_{n\in\mathbb{N}}$ ограничена. Пусть $\hat{v}$ — её предельная точка. Ясно, что $\hat{v} \neq v$. По (1) $\|u - \hat{v}\| = \lim_{n\to\infty} \|u_n - v_n\| = \rho(u, M)$. Так как M замкнуто, то $\hat{v} \in M$, т. е. обе точки $v, \hat{v}$ являются ближайшими к u — противоречие с условием, что M — чебышёвское множество.

Итак, метрическая проекция P непрерывна. Поэтому отображение g также непрерывно. Применим теорему Брауэра о неподвижной точке к выпуклому компакту B и отображению g: найдётся точка $z_0 \in B$ такая, что $g(z_0) = z_0$. Из определения $g(z_0)$ следует, что точка x лежит на отрезке, соединяющем z_0 с её ближайшим элементом Pz_0. При этом (как

легко следует из неравенства треугольника и того, что M — чебышёвское множество), ближайшим элементом к любой точке отрезка $[z_0; Pz_0]$ будет точка Pz_0. Но для x ближайшая точка $y = Px$, и она единственна. Значит, $Pz_0 = y$. Следовательно, для всех точек отрезка $[y; z_0]$ ближайшей точкой из M является точка y.

Применяя предыдущие рассуждения, проведённые для x, к точке z_0, мы ещё далее сдвинемся по лучу ℓ. В итоге для любой точки $w \in \ell$ точка y будет единственной ближайшей из M. Лемма 1 доказана. □

Лемма 2. *Всякое чебышёвское солнце в* $\mathbb{R}^2$ (в $\mathbb{R}^n$) *выпукло.*

Мы дадим два доказательства этого факта.

Доказательство. Способ I (аналитический) [8]. Пусть $x, y \in M$ и $z = \lambda x + (1 - \lambda)y \in [x; y]$, где $0 \leqslant \lambda \leqslant 1$. Поскольку M — чебышёвское солнце, то при $z \notin M$ каждая точка $z + \theta(z - Pz)$ для $\theta > 0$ имеет точку Pz своей ближайшей из M.

Поскольку Pz — ближайшая точка для $z + \theta(z - Pz)$, то

$$\|z + \theta(z - Pz) - x\|^2 \geqslant \|z + \theta(z - Pz) - Pz\|^2,$$

и, пользуясь свойствами скалярного произведения, мы получаем

$$\|z - x\|^2/\theta + 2\langle z - Pz, z - x\rangle \geqslant \|z - Pz\|^2/\theta + 2\langle z - Pz, z - Pz\rangle.$$

Устремляя $\theta \to \infty$ в этом неравенстве, мы получаем

$$\langle z - Pz, z - x\rangle \geqslant \|z - Pz\|^2.$$

Аналогично, подставляя в предыдущих рассуждениях y вместо x, имеем

$$\langle z - Pz, z - y\rangle \geqslant \|z - Pz\|^2.$$

Умножая два последних неравенства на λ и $(1 - \lambda)$ соответственно, складывая их и учитывая, что $z = \lambda x + (1 - \lambda)y$, находим, что

$$0 \geqslant \|z - Pz\|^2.$$

Итак, $\|z - Pz\| = 0$, откуда $z = Pz \in M$, т. е. M — выпукло.

Способ II (геометрический). Пусть $x \notin M$, $y = Px$. Посредством Γ обозначим касательную к кругу $B(x, \|x - y\|)$ в точке y. Без ограничения общности (ситуация не изменится при одновременном сдвиге или повороте рассматриваемых объектов), считаем, что прямая Γ совпадает

с осью абсцисс, а точка $x = (0, \xi)$ лежит на оси ординат и $\xi > 0$. Тогда $y = (0, 0)$.

По свойству солнечности любая точка w луча $\ell = \{\lambda x \,|\, \lambda > 0\}$ имеет своей ближайшей точку y. Поэтому прямая Γ будет касательной к любому кругу $B(w, \|w - y\|)$, $w \in \ell$.

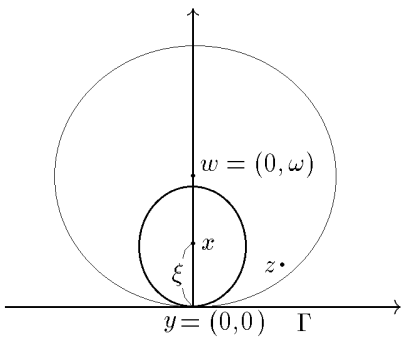


Рис. 7.

Рассмотрим $\Pi = \cup_{w \in \ell} \mathring{B}(w, \|w\|)$ — объединение открытых кругов с центрами на ℓ. Отметим, что $\mathring{B}(w, \|w\|) \cap M = \varnothing$ для $w \in \ell$. Пусть z — произвольная точка в верхней полуплоскости. Поскольку Γ — единственная касательная к любому кругу $B(w, \|w\|)$, $w \in \ell$, то интервал $(z; y)$ обязан пересекаться с внутренностью некоторого шара $\mathring{B}(x, \xi) \in \Pi$. Но тогда, если точка w имеет достаточно большую ординату, то точка z будет лежать в $\mathring{B}(w, \|w\|)$ (это выполнено, если $\omega > (\zeta_1^2 + \zeta_2^2)/(2\zeta_2)$, где $z = (\zeta_1, \zeta_2)$, $\zeta_2 > 0$, $w = (0, \omega)$). Следовательно, $z \in \Pi$ и верхняя полуплоскость совпадает с Π и не содержит точек из M.

Итак, любую точку $x \notin M$ можно отделить от множества M прямой, не содержащей x. Следовательно, M выпукло (действительно, если бы нашлись две точки $u, v \in M$ такие, что $x \in [u; v]$ и $x \notin M$, то точку x было бы невозможно отделить от M прямой). Поэтому чебышёвское солнце M выпукло. Лемма 2 доказана. □

Теперь *необходимость* в теореме Бунта немедленно следует из двух доказанных лемм: по лемме 1 каждое чебышёвское множество является чебышёвским солнцем, а по лемме 2 всякое солнце выпукло.

*** ***

Следующие два доказательства теоремы Бунта мы дадим в конечномерном *евклидовом пространстве* $\mathbb{R}^n$. Элемент этого пространства — любая совокупность $x = (\alpha_1, \alpha_2, \ldots, \alpha_n)$ из n действительных чисел. Эти числа $\alpha_1, \alpha_2, \ldots, \alpha_n$ называются координатами точки x. Действия сложения и умножения на число $\lambda \in \mathbb{R}$ производятся по следующим правилам:

$$x + y = (\alpha_1, \alpha_2, \ldots, \alpha_n) + (\beta_1, \beta_2, \ldots, \beta_n) = (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n),$$
$$\lambda x = \lambda(\alpha_1, \alpha_2, \ldots, \alpha_n) = (\lambda\alpha_1, \lambda\alpha_2, \ldots, \lambda\alpha_n).$$

Для любых $x = (\alpha_1, \ldots, \alpha_n)$, $y = (\beta_1, \ldots, \beta_n) \in \mathbb{R}^n$ определим *скалярное произведение* векторов x и y по формуле:

$$\langle x, y\rangle = \alpha_1\beta_1 + \ldots + \alpha_n\beta_n.$$

Это определение обобщает известную формулу выражения скалярного произведения в трёхмерном пространстве через координаты сомножителей в ортогональной системе координат. Легко видеть, что $\langle\cdot,\cdot\rangle$ удовлетворяет следующим требованиям: $\langle x, y\rangle = \langle y, x\rangle$, $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$, $\langle \lambda x, y\rangle = \lambda\langle x, y\rangle$, $\langle x, x\rangle > 0$ при $x \neq 0$ и $\langle x, x\rangle = 0$ при $x = 0$.

Длиной вектора x в евклидовом пространстве $\mathbb{R}^n$ называется число

$$\|x\| = \sqrt{\alpha_1^2 + \cdots + \alpha_n^2} = \sqrt{(x, x)}.$$

3. Доказательство Асплунда – Фиккена

Это доказательство, как и доказательство из следующего п. 4 (с помощью леммы об очистке), будет опираться на свойства множеств с единственной наиболее удалённой точкой и свойства инверсии единичной сферы, дающиеся ниже.

Пусть $x \in \mathbb{R}^n$ и $M \subset \mathbb{R}^n$. Точка $y_0 \in M$ называется *наиболее удалённой точкой* из M для x, если

$$\|x - y_0\| = \sup\{\|x - y\| \mid y \in M\}, \quad \text{т. е. } \|x - y_0\| \geqslant \|x - y\| \ \ \forall y \in M.$$

Скажем, что множество M обладает свойством *единственности наиболее удалённой точки*, если для любого $x \in \mathbb{R}^n$ в M существует в точности одна точка y_0, наиболее удалённая от x. Такие множества в каком-то смысле «обратны» к чебышёвским множествам. Отображение, сопоставляющее точке x её наиболее удалённую точку из M, обозначим посредством $F : \mathbb{R}^n \to M$. Итак,

$$F(x) \in M, \qquad \|x - F(x)\| = \sup\{\|x - y\| \mid y \in M\}.$$

Легко видеть, что примером множества со свойством единственности наиболее удалённой точки служит одноточечное множество. Другие примеры таких множеств *нельзя* построить, что показывает следующая важная теорема (см., напр., [8]).

Теорема (Йессен). *Только одноточечные множества в $\mathbb{R}^n$ обладают свойством единственности наиболее удалённой точки.*

Доказательство. Будем доказывать теорему только для замкнутых множеств. Пусть $M \subset \mathbb{R}^n$ — замкнутое множество со свойством единственности наиболее удалённой точки. Ясно, что M ограничено. Пусть $F : \mathbb{R}^n \to M$ — отображение, сопоставляющее точке x её наиболее удалённую точку из M. Покажем, что отображение F непрерывно и далее применим теорему Брауэра о неподвижной точке к произвольному шару $B(x_0, r_0)$, содержащему M.

Предположим, что отображение F *не* является непрерывным. Тогда найдётся последовательность $x_1, \ldots, x_k, \ldots$ точек из $\mathbb{R}^n$, сходящаяся к точке x, и такая, что последовательность $F(x_1), \ldots, F(x_k), \ldots$ не сходится к точке $F(x)$. Поскольку M ограничено, то найдётся подпоследовательность $x_{i_1}, \ldots, x_{i_m}, \ldots$ последовательности $x_1, \ldots, x_k, \ldots$ такая, что последовательность $F(x_{i_1}), \ldots, F(x_{i_m}), \ldots$ сходится к некоторой точке y, *отличной* от $F(x)$. Поскольку M замкнуто, то $y \in M$. Точки $F(x)$ и y наиболее удалены от x — противоречие с определением M. Поэтому отображение F непрерывно.

Сужение отображения F на непустое компактное выпуклое множество — произвольный содержащий M шар $B(x_0, r_0)$ — отображает его в себя. По теореме Брауэра существует неподвижная точка $y_0 \in B(x_0, r_0)$ отображения F, т. е. $F(y_0) = y_0$. Поэтому $y_0 \in M$. Итак, y_0 — наиболее удалённая точка из множества M для точки y_0, и, значит, множество M состоит из одной точки. □

Отображение $\sigma : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^n$, определяемое как $\sigma(x) = x/\|x\|^2$ называется *инверсией единичной сферы.* Оно непрерывно в любой точке x области определения, т. е. при $x \neq 0$. Следующая лемма описывает образ произвольного шара или сферы при инверсии σ.

Лемма 3. *Справедливы следующие соотношения:*
1) $\sigma\big(S(x,r)\big) = S\big(x/(\|x\|^2 - r^2), r/\big|\|x\|^2 - r^2\big|\big)$, $\|x\| \neq r$,
2) $\sigma\big(S(x,\|x\|)\big) = \{z \mid \langle z, x\rangle = 1/2\}$,
3) $\sigma\big(B(x,r)\big) = B\big(x/(\|x\|^2 - r^2), r/(\|x\|^2 - r^2)\big)$, *если* $\|x\| > r$,
4) $\sigma\big(B(x,r) \setminus \{0\}\big) = \mathbb{R}^n \setminus \mathring{B}\big(x/(\|x\|^2 - r^2), r/\big|\|x\|^2 - r^2\big|\big)$, *если* $\|x\| < r$,
5) $\sigma\big(B(x,\|x\|) \setminus \{0\}\big) = \{z \in \mathbb{R}^n \mid \langle z, x\rangle \geqslant 1/2\}$.

Переходим собственно к *доказательству теоремы Бунта.* Предположим, что в $\mathbb{R}^n$ существует невыпуклое чебышёвское множество M. Без ограничения общности считаем, что $0 \notin M$, $0 \in \operatorname{conv} M$ (через $\operatorname{conv} M$ мы обозначаем *выпуклую оболочку* множества M, т. е. минимальное выпуклое множество, содержащее M). Например, выпуклая оболочка окружности — это круг, а выпуклая оболочка двух различных точек — это

отрезок, их соединяющий. Отметим, что множество M выпукло тогда и только тогда, когда оно совпадает со своей выпуклой оболочкой $\operatorname{conv} M$.

Инверсия $\sigma : x \mapsto x/\|x\|^2$ $(x \neq 0)$ единичной сферы переводит M в ограниченное множество

$$G = \sigma(M) = \{x/\|x\|^2 \,\big|\, x \in M\}.$$

Каждый шар $B(x, r)$, содержащий G, обязан содержать начало координат в своей внутренности, т. е. должно выполняться неравенство $\|x\| < r$ (иначе множество M содержалось бы в замкнутом полупространстве, *не* содержащем начала координат, что противоречит условию $0 \in \operatorname{conv} M$).

Для точки $x \in \mathbb{R}^n$ найдётся такой минимальный шар $B(x, t(x))$, содержащий G, что $G \not\subset B(x, r)$ для всех $r < t(x)$. Под действием инверсии шар $B(x, t(x))$ переходит в дополнение $\mathbb{R}^n \setminus V$ к открытому шару

$$V = \mathring{B}\left(x/(\|x\|^2 - t^2(x)), \frac{t(x)}{t^2(x) - \|x\|^2}\right),$$

который, в свою очередь, является максимальным шаром с центром $x/\left(\|x\|^2 - t^2(x)\right)$ таким, что замыкание его дополнения содержит M. Поскольку M — чебышёвское множество, этот шар пересекает M в единственной точке $P(x/(\|x\|^2 - t^2(x)))$. Следовательно, функция

$$x \mapsto q(x) = \frac{P(x/(\|x\|^2 - t^2(x)))}{\|P(x/(\|x\|^2 - t^2(x)))\|},$$

действующая из $\mathbb{R}^n$ в G, обладает свойством

$$\|x - y\| < \|x - q(x)\|, \quad \text{если } y \in G \text{ и } y \neq q(x).$$

Итак, G обладает свойством единственности наиболее удалённой точки. Поэтому по теореме Йессена оно, а, следовательно, и множество M одноточечно. Поэтому исходное предположение $0 \notin M$, $0 \in \operatorname{conv} M$ не может выполняться для такого одноточечного M. Итак, M — выпукло. □

4. Доказательство С.В. Конягина при помощи леммы об очистке

Часто является полезной следующая лемма «об очистке» для субдифференциалов (см. [9], мы приводим её упрощённый вариант). В ней уменьшается количество перебираемых элементов, происходит «очистка» от «ненужных» точек.

Лемма (об очистке, [7]). *Пусть T — компактное множество в $\mathbb{R}^n$, $g : T \times \mathbb{R}^n \to \mathbb{R}^n$ — функция, выпуклая по x при всяком $t \in T$ и непрерывная по совокупности переменных (t, x). Положим $\varphi(x) = \max_{t\in T} g(t, x)$. Тогда, если $\hat{x}$ — точка минимума φ, то найдутся положительные числа $\alpha_1, \dots, \alpha_r$, $\alpha_1 + \dots + \alpha_r = 1$, $r \leqslant n + 1$ и точки $y_1, \dots, y_r$, $y_i \in \partial g(\tau_i, \cdot)(\hat{x})$ (где $\tau_i \in T_0(\hat{x}) = \{t \in T \mid g(t, \hat{x}) = \varphi(\hat{x})\}$ и $\partial g(\tau_i, \cdot)(\hat{x}) = \{y \in \mathbb{R}^n \mid g(\tau_i, x) - g(\tau_i, \hat{x}) \geqslant \langle x - \hat{x}, y_i\rangle, \forall x \in \mathbb{R}^n\}$, $i = 1, \dots, r$), такие, что $0 = \sum_{i=1}^{r} \alpha_i y_i$.*

Доказательство теоремы. Пусть M — невыпуклое чебышёвское множество в $\mathbb{R}^n$ и пусть $G = \{x/\|x\|^2 \mid x \in M\}$ — его образ при инверсии $x \mapsto x/\|x\|^2$. Множество $G \subset \mathbb{R}^n$ ограничено и замкнуто, т. е. компактно. Выше мы показали также, что G обладает свойством единственности наиболее удалённого элемента. Применим лемму об очистке для множества $T = G$, функции $g(t, x) = \|t - x\|$, $t \in G$. Пусть $\hat{x}$ — точка минимума функции φ и пусть числа α_i, точки τ_i, y_i $(i = 1, \dots, n)$ и множество T_0 такие, как в лемме.

Предположим, что $r > 1$. Тогда по определению T_0 точки $\tau_1, \dots, \tau_r$ являются наиболее удалёнными точками из множества G для точки $\hat{x}$ — противоречие, поскольку G — множество со свойством *единственности* наиболее удалённой точки.

Пусть теперь $r = 1$. Тогда $T_0 = \{\tau_1\}$ и по лемме об очистке $0 = \alpha_1 y_1$. Поскольку $\alpha_1 = 1$, то $y_1 = 0$. С другой стороны, $y_1 \in \partial g(\tau_1, \cdot)(\hat{x})$. Это по определению означает, что $\|\tau_1 - x\| - \|\tau_1 - \hat{x}\| \geqslant 0$ для всех $x \in \mathbb{R}^n$, что влечёт $\|\tau_1 - \hat{x}\| = 0$, т. е. $\tau_1 = \hat{x}$. Но τ_1 — наиболее удалённая точка из G для $\hat{x}$. Поэтому по теореме Йессена множество G, а, следовательно, и множество M, состоят из одной точки, откуда M не может быть невыпуклым. Противоречие с начальным предположением. □

5. Доказательство Л. П. Власова

Лемма 4. *Пусть M — чебышёвское множество*[3)]. *Тогда для любой точки $x \notin M$ и для любой последовательности $\{x_n\}_{n\in\mathbb{N}}$, $x \in (x_n; Px)$ $(n \in \mathbb{N})$, $x_n \to x$, выполнено*

$$\frac{\rho(x_n, M) - \rho(x, M)}{\|x_n - x\|} \longrightarrow 1. \tag{2}$$

[3)] Лемма верна и в более общем случае банаховых пространств для чебышёвских множеств с непрерывной метрической проекцией.

Доказательство леммы 4. Пусть $x \in \mathbb{R}^n \setminus M$, $y = Px$ и пусть последовательность $\{x_n\}_{n\in\mathbb{N}}$ такова, что $x \in (x_n; y)$ и $x_n \to x$. Положим $y_n = Px_n$.

Если $y = y_n$ для некоторого $n \in \mathbb{N}$, то для любой точки $\xi \in [y_n; x]$ выполнено $P\xi = y$, что и требуется в (2).

Предположим теперь, что $y \neq y_n$ для всех n. В плоскости, определяемой точками x, x_n, y, y_n, найдём точку z пересечения прямой $x_n z$, параллельной xy_n, с прямой yy_n. Из подобия треугольников $\triangle xyy_n$ и $\triangle x_n yz$ находим

$$\|z-y_n\|/\|x_n-x\| = \|y_n-x_n\|/\|x-y\|, \quad \|x_n-y\|/\|x_n-z\| = \|x-y\|/\|x-y_n\|. \tag{3}$$

Соотношения $y = Px$ и $y_n = Px_n$ дают соответственно

$$\|x - y\| \leqslant \|x - y_n\|, \quad \|x_n - y_n\| \leqslant \|x_n - y\|. \tag{4}$$

Из (3) и (4) вытекает

$$\|x_n - y\| \leqslant \|x_n - z\|. \tag{5}$$

Применяя последовательно равенство $\|x_n - y\| = \|x_n - x\| + \|x - y\|$, неравенства (4) и (5) и $\|x_n - z\| \leqslant \|x_n - y_n\| + \|y_n - z\|$, а затем неравенство (3), получим

$$\begin{aligned} 0 \leqslant 1 - \frac{\rho(x_n, M) - \rho(x, M)}{\|x_n - x\|} &= 1 - \frac{\|x_n - y_n\| - \|x - y\|}{\|x_n - x\|} = \\ = \frac{\|x_n - y\| - \|x_n - y_n\|}{\|x_n - x\|} &\leqslant \frac{\|x_n - z\| - \|x_n - y_n\|}{\|x_n - x\|} \leqslant \\ &\leqslant \frac{\|y_n - z\|}{\|x_n - x\|} = \frac{\|y_n - y\|}{\|x - y\|} = \frac{\|y_n - y\|}{\rho(x, M)} \longrightarrow 0, \end{aligned}$$

поскольку $y_n \to y$ по доказанной в лемме 1 непрерывности метрической проекции. □

Доказательство теоремы Бунта. Пусть M — чебышёвское множество. По лемме 4 оно удовлетворяет условию (2). Пусть $x_0 \in \mathbb{R}^n \setminus M$, $\delta > 0$ и $R = \rho(x_0, M)$. Для $\xi \in \mathbb{R}^n$ положим $f(\xi) = (1-2\delta)\|\xi - x_0\| - \rho(\xi, M)$. Непрерывная функция f достигает своего минимума на замкнутом ограниченном множестве $B(x_0, R)$ в некоторой точке x. Покажем, что $x \in S(x_0, R)$.

Предположим противное, т. е. $\|x - x_0\| < R$. Тогда по свойству (2) найдётся точка $y \in \mathring{B}(x, R)$ такая, что $\rho(y, M) - \rho(x, M) > (1 - \delta)\|x - y\|$.

Тогда

$$\begin{aligned} f(y) &= (1-2\delta)\|y-x_0\| - \rho(y,M) < \\ &< (1-2\delta)(\|y-x\| + \|x-x_0\|) - \rho(x,M) - (1-\delta)\|x-y\| = \\ &= (1-2\delta)\|x_0-x\| - \rho(x,M) - \delta\|x-y\| = f(x) - \delta\|x-y\|. \end{aligned}$$

Полученное неравенство противоречит неравенству $f(x) \leqslant f(y)$. Следовательно, предположение $x \in \mathring{B}(x_0, R)$ было неверно и, окончательно, $\|x_0 - x\| = R$.

Поскольку $f(x) \leqslant f(x_0)$, то

$$\rho(x,M) \geqslant \rho(x_0,M) + (1-2\delta)R.$$

Пусть $\delta_n \to 0$. Функция $f_n(\xi) = (1-2\delta_n)\|\xi - x_0\| - \rho(\xi, M)$ достигает минимума на $B(x_0, R)$ в точке x_n. По доказанному выше $\|x_0 - x_n\| = R$ $(n = 1, 2, \dots)$ и

$$\rho(x_n,M) \geqslant \rho(x_0,M) + (1-2\delta_n)R. \tag{6}$$

Положим $y_0 = Px_0$, $s = (x_0 - y_0)/\|x_0 - y_0\|$ и $s_n = (x_n - x_0)/\|x_n - x_0\|$. Поскольку $\|x_n - y_0\| \geqslant \|x_n - y_n\|$, то из (6) следует неравенство $\|x_n - y_0\| \geqslant \|x_0 - y_0\| + R - 2R\delta_n$, что даёт $\|s + s_n\| \geqslant 2 - 2\delta_n$. Для векторов s, s_n запишем равенство параллелограмма:

$$\|s + s_n\|^2 + \|s - s_n\|^2 = 2(\|s\|^2 + \|s_n\|^2) = 4. \tag{7}$$

Поскольку $\delta_n \to 0$, то из (7) с учётом предыдущего неравенства получаем $\|s - s_n\| \to 0$ при $n \to \infty$. Из этого сразу следует, что $x_n \to \hat{x} = y_0 - 2x_0$ при $n \to \infty$ (точка $\hat{x}$ диаметрально противоположна точке y_0 на сфере $S(x_0, R)$). Переходя к пределу при $n \to \infty$ в (6) получим

$$\rho(\hat{x},M) \geqslant \rho(x_0,M) + R. \tag{8}$$

С другой стороны,

$$\rho(\hat{x},M) \leqslant \|\hat{x} - y_0\| \leqslant \|\hat{x} - x_0\| + \rho(x_0,M) = R + \rho(x_0,M). \tag{9}$$

Из (8) и (9) следует

$$\rho(\hat{x},M) = \|\hat{x} - y_0\| = 2R,$$

что даёт $\rho(\hat{x}, M) = 2R$, откуда $P\hat{x} = y_0$.

Итак, для точки $x_0 \notin M$ мы показали, что точка $\hat{x}$ луча ℓ, начинающегося в $y_0 = Px_0$ и проходящего через x_0, имеет точку y_0 своей ближайшей. Напомним, что x_0 — середина отрезка $[y_0; \hat{x}]$. Следовательно, и

любая другая точка отрезка $[y_0; \hat{x}]$ имеет y_0 своей ближайшей. Применяя предыдущие рассуждения, проведённые для x_0, к точке $\hat{x}$, мы ещё далее сдвинемся по лучу ℓ. В итоге, для любой точки $w \in \ell$ точка y_0 будет единственной ближайшей из M. Значит, M — чебышёвское солнце, и по лемме 2 оно выпукло. Теорема доказана. □

Некоторые обобщения

В бесконечномерных линейных нормированных пространствах о выпуклости чебышёвских множеств известно немного. Например, до сих пор не известно, выпукло ли произвольное чебышёвское множество в бесконечномерном гильбертовом пространстве.

Проблема 1. (Н. В. Ефимов, С. Б. Стечкин, В. Кли) *Доказать или опровергнуть, что всякое чебышёвское множество в бесконечномерном гильбертовом пространстве выпукло.*

Напомним [5], что линейное нормированное пространство, на котором введено скалярное произведение $\langle \cdot, \cdot \rangle$, называется предгильбертовым пространством. Гильбертовым пространством называется предгильбертово пространство, полное относительно нормы $\|x\| = \sqrt{\langle x, x \rangle}$. Конечномерное гильбертово пространство — это пространство $\mathbb{R}^n$.

Требование полноты предгильбертова пространства в проблеме 1 важно, как показывает пример *невыпуклого* чебышёвского множества, построенный Джонсономw (см. [2]) в неполном предгильбертовом пространстве ℓ_0^2 последовательностей, оканчивающихся нулями. Также неизвестно, есть ли бесконечномерное пространство, в котором всякое чебышёвское множество выпукло.

При некоторых ограничениях на структуру чебышёвского множества (различные виды компактности) или при ограничении на свойства метрической проекции (непрерывность в каком-либо смысле) удаётся доказать его выпуклость.

Напомним, что множество называется *ограниченно компактным*, если его пересечение с любым замкнутым шаром компактно. Понятно, что компактное множество является ограниченно компактным.

Теорема (Л. П. Власов). *Пусть X — произвольное банахово пространство. Тогда всякое ограниченно компактное чебышёвское множество в X является чебышёвским солнцем, а в гладком X — выпукло.*

Доказательство почти дословно повторяет приведённое в п. 2, но вместо конечномерной теоремы Брауэра применяется её обобщение на случай бесконечномерных пространств (теорема Шаудера [6]): *всякое непрерывное отображение выпуклого замкнутого множества в свою компактную часть имеет неподвижную точку.*

Рассуждения в последнем доказательстве могут быть обобщены на случай чебышёвских множеств с непрерывной метрической проекцией в банаховых пространствах, удовлетворяющих следующему свойству

$$s, s_n \in S(0,1),\ \|s + s_n\| \to 2 \text{ при } n \to \infty \Longrightarrow\ s_n \to s \text{ при } n \to \infty. \qquad (10)$$

Если X — гильбертово пространство (например, $\mathbb{R}^n$), то X удовлетворяет условию (10) в силу равенства параллелограмма.

Приведённое доказательство остаётся без изменения, за исключением следующего момента. В $\mathbb{R}^n$ шар $B(x_0, R)$ компактен, и поэтому непрерывная функция f достигает своего минимума на $B(x_0, R)$. В бесконечномерном случае этот факт может быть неверен, но по вариационному принципу Экланда функция f достигает «почти минимума» на $B(x_0, R)$: для любого $\varepsilon > 0$ найдётся $\hat{x} \in B(x_0, R)$ такой, что для любого $y \in B(x_0, R)$ выполнено $f(y) \geqslant f(x) - \varepsilon\|x - y\|$. Далее применяем предыдущие рассуждения.

Несмотря на «очень хорошее» с точки зрения геометрии устройство чебышёвских множеств в n-мерных евклидовых пространствах, в произвольных *конечномерных* линейных нормированных пространствах X_n о выпуклости чебышёвских множеств известно далеко не всё. Пространство называется *гладким*, если в каждой точке единичной сферы касательная гиперплоскость к сфере единственна. К примеру, пространства $\ell^p(n)$, $1 < p < \infty$, с нормой $\|x\|_p = (\alpha_1^p + \ldots + \alpha_n^p)^{1/p}$, $x = (\alpha_1, \ldots, x_n)$, являются гладкими пространствами, а пространство X_2 с max-нормой («квадрат») $\|x\|_\infty = \max\{|\alpha_1|, |\alpha_2|\}$ — негладко.

Моцкин показал, что гладкость двумерного пространства X_2 необходима и достаточна для выпуклости всякого чебышёвского множества в X_2. Рассуждая, как в лемме 2 (способ II), несложно показать, что в гладких пространствах всякое чебышёвское множество выпукло. Однако, как независимо показали В.И. Бердышев и А. Брондстед, всякое чебышёвское множество выпукло и в некоторых *негладких* пространствах X_n для всех $n \geqslant 3$.

Сформулируем теорему, характеризующую пространства X_n ($n \leqslant 4$), в которых всякое чебышёвское множество выпукло. При $n = 2$ она была доказана Моцкиным, при $n = 3$ — Бердышевым и Брондстедом, а при $n = 4$ — А.Л. Брауном. Напомним, что точка $s \in S(0, 1)$ называется *достижимой*, если найдётся касательная гиперплоскость H к сфере в точке s такая, что $H \cap S(0, 1) = \{s\}$; точка s называется точкой *гладкости*, если в ней касательная гиперплоскость к сфере единственна.

Теорема. *В конечномерном линейном нормированном пространстве X_n ($n \leqslant 4$) всякое чебышёвское множество выпукло тогда и только тогда, когда всякая достижимая точка сферы является точкой гладкости.*

Для $n \geqslant 5$ вопрос о характеризации пространств X_n, в которых всякое чебышёвское множество выпукло, остаётся открытым. Есть гипотеза, что характеризация, данная в предыдущей теореме, верна и для любого $n \geqslant 5$. Отметим очень интересный результат И. Г. Царькова [9], который охарактеризовал *все* конечномерные пространства X_n, в которых всякое *ограниченное* чебышёвское множество выпукло.

Список литературы

[1] *Asplund E.* Čebyšev sets in Hilbert space // Trans. Amer. Math. Soc. V. 144, 1969. P. 235–240.

[2] *Балаганский В.С., Власов Л.П.* Проблема выпуклости чебышёвских множеств // УМН. Т. 51, вып. 6, 1996. С. 125–188.

[3] *Власов Л.П.* Аппроксимативные свойства множеств в линейных нормированных пространствах // УМН. Т. 28, вып. 6, 1973. С. 3–66.

[4] *Лейхтвейс К.* Выпуклые множества. М.: Наука, 1985.

[5] *Люстерник Л.А., Соболев С.А.* Элементы функционального анализа. М.: Наука, 1965.

[6] *Smart D.R.* Fixed Point Theorems. *Cambridge Tracts in Mathematics*, 66. Cambridge, 1974.

[7] *Тихомиров В.М.* Некоторые вопросы теории приближений. М.: Изд-во МГУ, 1976.

[8] *Webster R.* Convexity. Oxford, 1994.

[9] *Царьков И.Г.* Ограниченные чебышёвские множества в банаховых пространствах // Матем. заметки. Т. 36, № 1, 1984. С. 73–87.

Ширина многоугольника

М. Л. Гервер*

Если вырезанный из картона выпуклый многоугольник можно, непрерывно двигая по столу, протащить между двумя гвоздями, вбитыми в стол на расстоянии 1, то тогда его можно протащить и между гвоздями, вбитыми на расстоянии $r > 1$. Верно ли то же самое и для любого невыпуклого многоугольника, — например, для такого, как на рис. 1?

Библиография: 2 названия.

При дополнительных ограничениях (которым многоугольник M на рис. 1 не удовлетворяет) утвердительный ответ на вопрос из аннотации получили в [1] J. E. Goodman, J. Pach и C. K. Yap. Методы и результаты работы [1] послужили основой для построений настоящей статьи.

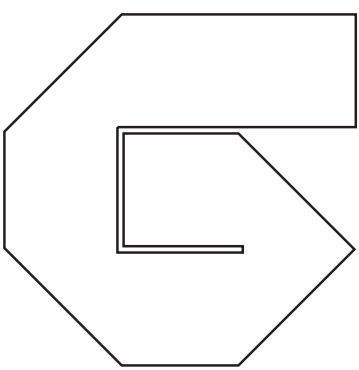


Рис. 1

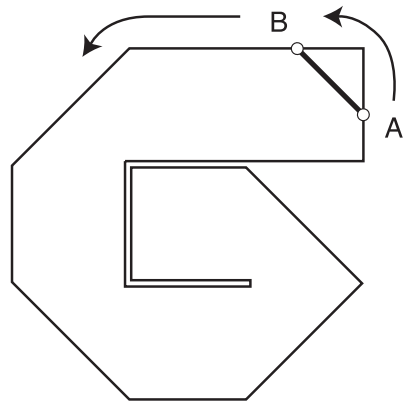


Рис. 2

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант 96-01-01852).

1. Теорема двойственности

Минимальное расстояние w_M между двумя точками в $\mathbb{R}^2$, позволяющее протащить плоский многоугольник M между ними, назовем *шириной* M (можно доказать, что для выпуклого многоугольника данное определение совпадает со стандартным: w_M — наименьшее расстояние между параллельными прямыми, опорными к M).

Из новых результатов, полученных в статье, отметим следующий.

Пусть A и B — две точки на границе P многоугольника M (рис. 2). Ясно, что если длина d отрезка AB не слишком велика, то его можно непрерывно протащить вдоль P — так, что его концы A и B всё время будут находиться на P и отрезок AB, сделав полный оборот вокруг M, вернется в исходное положение.

Спрашивается: каково максимальное значение d_M длины d, допускающее такое перемещение концов отрезка по границе M?

Оказывается, d_M в точности равно ширине w_M многоугольника M.

Чтобы доказать эту теорему двойственности и решить задачу, сформулированную в аннотации, рассмотрим ряд подготовительных, вспомогательных задач.

2. Возы и машины. Фазовое пространство

Разобраться с задачами о ширине многоугольника нам, в первую очередь, поможет прием, используемый в [2] при решении следующей задачи Н. Н. Константинова.

Две непересекающиеся дороги ведут из города C_1 в город C_2. Известно, что две машины, связанные веревкой длины $2L$, могут проехать из C_1 в C_2 (по разным дорогам), не порвав веревки. Смогут ли разминуться два круглых воза радиуса $R > L$, если их центры движутся по тем же дорогам навстречу друг другу?

Решение. Вот — решение этой задачи, приведенное в [2, с. 8,9]. Рассмотрим квадрат

$$K = \{(x_1, x_2) \,|\, 0 \leqslant x_i \leqslant 1,\ i = 1, 2\}.$$

Положение двух экипажей (один — на 1-й дороге, другой — на 2-й) можно характеризовать точкой квадрата K: достаточно обозначить через x_i долю расстояния от C_1 до C_2 по i-той дороге, заключённую между C_1 и находящимся на этой дороге экипажем.

Всевозможным положениям экипажей соответствуют всевозможные точки квадрата K. Этот квадрат называется *фазовым пространством*,

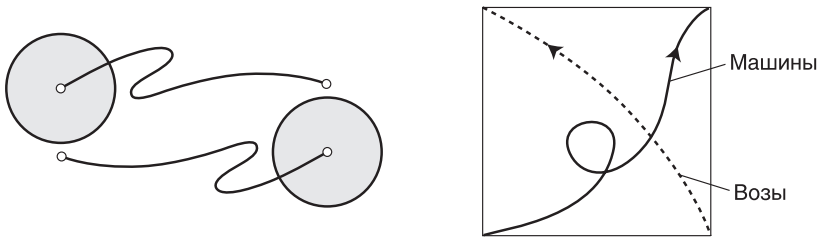


Рис. 3

а его точки — *фазовыми точками*. Таким образом, каждая фазовая точка соответствует определенному положению пары экипажей, а всякое движение экипажей изображается движением фазовой точки в фазовом пространстве.

Например, начальное положение машин (в городе C_1) соответствует левому нижнему углу квадрата K ($x_1 = x_2 = 0$), а движение машин из C_1 в C_2 изображается кривой, ведущей в противоположный угол.

Точно так же начальное положение возов (один — на 1-й дороге на выезде из C_2, другой — на 2-й дороге на выезде из C_1, рис. 3) соответствует правому нижнему углу квадрата K ($x_1 = 1$, $x_2 = 0$), а движение возов изображается кривой, ведущей в противоположный угол квадрата.

Но всякие две кривые в квадрате, соединяющие разные пары противоположных вершин, пересекаются. Поэтому, как бы ни двигались возы, наступит момент, когда пара возов займёт положение, в котором была в некоторый момент времени пара машин. В этот момент расстояние между центрами возов будет меньше $2R$. Итак, возам не удастся разминуться.

3. Многоугольник с ручкой и его ширина

Пусть полупрямая H (рис. 4 на след. стр.) имеет единственную общую точку с границей P многоугольника M. Объединение P и H обозначим через P_H и будем называть *многоугольником с ручкой* (взамен более точных, но тяжеловесных *контур* или *граница многоугольника с ручкой*), а саму полупрямую H будем называть *ручкой*.

С изучаемыми вопросами тесно связаны следующие задачи.

1. *Пусть A_0 и B_0 — две произвольные точки на ручке H. Доказать, что существует пара непрерывных отображений*

$$a:\ t \in [0;1] \to A = a(t) \in P_H, \quad b:\ t \in [0;1] \to B = b(t) \in P_H,$$

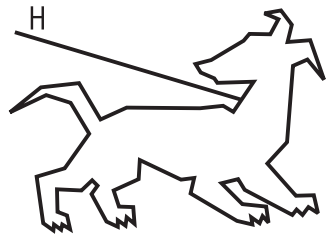


Рис. 4

обладающая следующими свойствами:
– *при* $t = 0$ *точки* A *и* B *занимают положения* A_0 *и* B_0*;*
– *при изменении* t *от* 0 *до* 1 *расстояние* $|AB|$ *остается неизменным;*
– *при изменении* t *от* 0 *до* 1 *по крайней мере одна из точек* A *и* B *обходит* P (*т.е. двигаясь непрерывно, но не обязательно монотонно — то по часовой, то против часовой стрелки — делает полный оборот вокруг* M)*;*
– *при* $t = 1$ *отрезок* AB *совпадает с отрезком* A_0B_0 — *так что либо*

$$A_1 = a(1) = A_0, \quad B_1 = b(1) = B_0, \tag{1}$$

либо

$$A_1 = a(1) = B_0, \quad B_1 = b(1) = A_0. \tag{2}$$

2. *Доказать, что условия* (1) *или* (2) *выполняются в зависимости от того, какова длина отрезка* A_0B_0*: существует такое* w*, что при* $|A_0B_0| > w$ *выполняется* (1)*, а при* $|A_0B_0| < w$ *выполняется* (2)*.*

3. *Верно ли, что при* $|A_0B_0| = w$ *существуют обе пары отображений: и такая, что выполняется* (1)*, и такая, что выполняется* (2)*?*

Определение. Число w назовем *шириной* многоугольника с ручкой P_H.

Дальнейшие построения оправдают выбор этого термина, а пока возникает вопрос:

4. *Зависит ли ширина* P_H *от расположения ручки* H*? Подробнее: пусть* H^* *— любая полупрямая, имеющая единственную общую точку с границей* P *многоугольника* M*. Объединим* P *с* H^* *и ширину полученного многоугольника с ручкой* P_{H^*} *обозначим через* w^**. Равны ли между собой* w *и* w^**?*

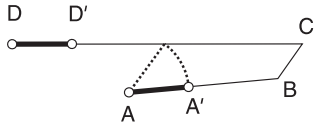


Рис. 5

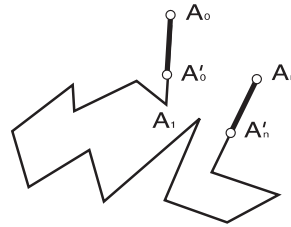


Рис. 6

4. Движение отрезка по ломаной

Прежде чем перейти к решению задач 1–4, обсудим следующее утверждение.

5. *Пусть $ABCD$ — несамопересекающаяся трёхзвенная ломаная линия, причем звенья AB и CD имеют (каждое) длину больше 1. Пусть A' и D' — такие точки на звеньях AB и CD, что $|AA'| = |D'D| = 1$. Тогда отрезок длины 1 можно так переместить из положения AA' в положение $D'D$, что концы отрезка будут непрерывно двигаться по ломаной $ABCD$.*

Из рис. 5 ясно, что утверждение 5 неверно.

6. *Станет ли утверждение 5 верным не только для трёхзвенной ломаной, но даже для n–звенной ломаной $A_0A_1\ldots A_n$, если дополнительно потребовать, чтобы все точки ломаной, лежащие на расстоянии $\leqslant 1$ от A_0, принадлежали её начальному звену A_0A_1, а все точки ломаной, лежащие на расстоянии $\leqslant 1$ от A_n, принадлежали её конечному звену?*

Решение задачи 6 (см. разд. 5) поможет лучше понять задачи 1–4 и, в частности, объяснить, *зачем к многоугольнику приделана ручка.*

5. Решение задачи о движении отрезка по ломаной. Граф Г

Пусть *все точки несамопересекающейся n-звенной ломаной $P = A_0A_1\ldots A_n$, лежащие на расстоянии $\leqslant 1$ от A_0, принадлежат её начальному звену A_0A_1, а все точки ломаной P, лежащие на расстоянии $\leqslant 1$ от A_n, принадлежат её конечному звену.* Обозначим через A'_0 и через A'_n точки P на расстоянии 1 от A_0 и от A_n и, следуя [1], докажем, что отрезок длины 1 можно непрерывно переместить из положения $A_0A'_0$ в положение A'_nA_n, двигая концы отрезка по ломаной P (рис. 6).

Пусть L — длина ломаной P и пусть AB — произвольный отрезок длины 1 с концами на P. Длины ломаных A_0A и A_0B (измеренные вдоль P)

обозначим через L_1 и L_2 и положим (сравним с разд. 2)

$$x_1 = L_1/L, \quad x_2 = L_2/L. \tag{3}$$

Поступив так с *каждым* отрезком длины 1 с концами на P, получим некоторое множество Γ точек с координатами (3) в *фазовом пространстве*, т.е. в квадрате

$$K = \{(x_1, x_2) \,|\, 0 \leqslant x_i \leqslant 1,\ i = 1, 2\}.$$

В частности, отрезкам A_0A_0' и $A_n'A_n$ соответствуют фазовые точки

$$F_0 = (0, 1/L) \quad \text{и} \quad F_1 = (1 - 1/L, 1). \tag{4}$$

Уточнение. Вместе с каждой точкой $X = (x_1, x_2)$ множество Γ, очевидно, содержит точку $X' = (x_2, x_1)$, симметричную X относительно прямой $x_1 = x_2$ (если фазовая точка X соответствует отрезку AB, то X' соответствует отрезку BA). Исключим из Γ все точки, у которых $x_1 > x_2$, т. е. при определении Γ будем рассматривать только такие отрезки AB, у которых (при движении вдоль P) A ближе к A_0, чем B.

Нетрудно проверить, что множество Γ представляет собой граф, рёбра которого являются либо дугами эллипсов, либо прямолинейными отрезками (отрезки получаются, когда концы AB принадлежат или одному и тому же звену ломаной $A_0A_1 \ldots A_n$, или параллельным звеньям, а дуги — когда концы AB принадлежат непараллельным звеньям). Степени вершин Γ описаны при доказательстве следующей теоремы (леммы 1, 2).

Теорема 1. *Фазовые точки F_0 и F_1 принадлежат одной и той же связной компоненте графа Γ, так что отрезок длины 1 можно непрерывно переместить из положения A_0A_0' в положение $A_n'A_n$, двигая концы отрезка по ломаной P.*

Доказательство. Допустим (потом мы освободимся от этого допущения), что

ломаная P не содержит ни одной пары параллельных звеньев, расстояние между которыми в точности равно 1. (5)

Докажем следующее утверждение.

Лемма 1. *При условии* (5) *все вершины графа Γ, кроме «концевых» вершин* (4), *имеют чётную степень:* $0, 2$ *или* 4, *и лишь вершины F_0 и F_1 имеют степень* 1.

Вместе с леммой 1 будет доказана и теорема 1 при условии (5): выйдя из F_0, будем двигаться по рёбрам графа Г, не проходя ни по одному ребру дважды (пока это возможно); это движение не может закончиться ни в какой вершине чётной степени (войдя в неё, мы можем выйти); тем самым мы дойдем до F_1.

Доказательство леммы 1. Пусть отрезок AB длины 1 с вершинами на ломаной P не совпадает ни с A_0A_0', ни с $A_n'A_n$. Тогда

$$A \neq A_0, \quad B \neq A_n \tag{6}$$

(здесь используется ограничение, наложенное на ломаную $P = A_0A_1 \dots A_n$ в начале параграфа: все точки ломаной P, лежащие на расстоянии $\leqslant 1$ от A_0, принадлежат её начальному звену, а все точки ломаной P, лежащие на расстоянии $\leqslant 1$ от A_n, принадлежат её конечному звену).

Вследствие (6) AB образует с P два угла в точке A (назовем их α_1, α_2) и два угла в точке B (назовем их β_1, β_2); при этом возможны следующие случаи:

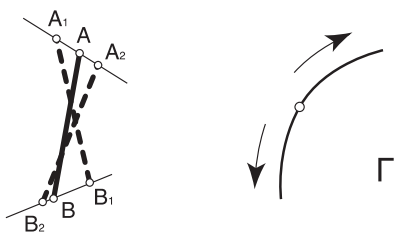


Рис. 7 a)

a) в точности один из углов α_1, α_2 — острый и в точности один из углов β_1, β_2 — острый; в этом (общем) случае оба конца AB могут одновременно двигаться по ломаной P в двух направлениях (рис. 7a);

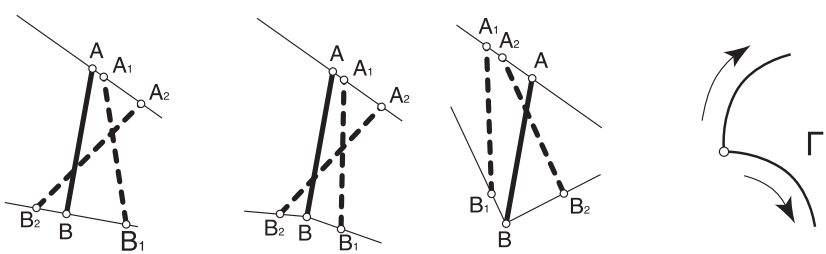


Рис. 7 b)

b) либо один, либо три из углов α_1, α_2, β_1, β_2 — острые; в этом случае один конец отрезка AB может двигаться по ломаной P в двух направлениях, а другой конец AB — в одном направлении (соответствующая вершина графа Γ имеет степень 2, рис. 7b);

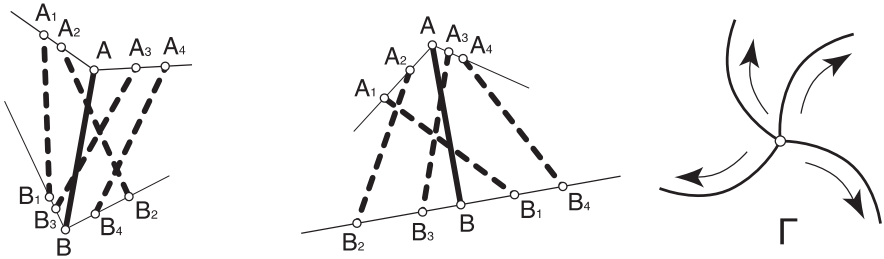
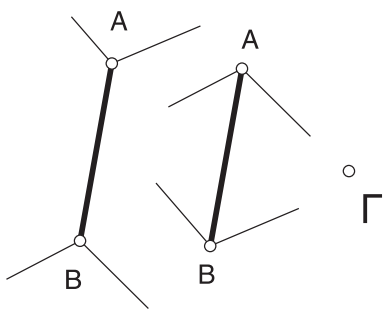


Рис. 7 c)

c) оба угла α_i — острые и ни один из углов β_i не является острым (или наоборот, оба угла β_i — острые и ни один из углов α_i не является острым); в этом случае каждый из концов отрезка AB может двигаться по ломаной P в двух направлениях, так что всего имеется четыре варианта движения (соответствующая вершина графа Γ имеет степень 4, рис. 7c);



k \ j	0	1	2
0	0	2	4
1	2	2	2
2	4	2	0

Рис. 7 d)

d) все углы α_1, α_2, β_1 и β_2 — острые, либо ни один из этих углов не является острым; из этого положения отрезок AB никуда не может сдвинуться (граф Γ имеет изолированную вершину, рис. 7d).

Случаи a–d исчерпывают 9 возможностей (j, k), $0 \leqslant j, k \leqslant 2$, где j обозначает число острых углов среди углов α_i, а k — число острых углов среди углов β_i, $i = 1, 2$. Лемма 1 — а с ней и теорема 1 при условии (5) — доказана.

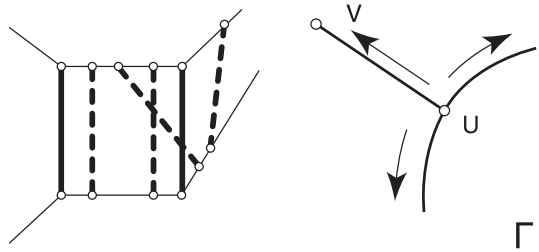


Рис. 7 е)

В общем случае — без условия (5) — граф Г, кроме «концевых» вершин F_i степени 1, может иметь и другие, «неконцевые», вершины нечётной степени.

Верна, однако, следующая лемма:

Лемма 2. *Все неконцевые вершины графа* Г *нечётной степени можно разбить на пары вершин* (U, V), *соединённых в* Г *прямолинейными отрезками — рёбрами* UV.

Докажем лемму 2, а потом выведем из неё теорему 1.

Доказательство леммы 2. Граф Г может иметь *неконцевые* вершины нечётной степени только при нарушении (5), т.е. при условии, что ломаная P содержит параллельные звенья, расстояние между которыми в точности равно 1 (рис. 7е).

Каждой такой паре звеньев соответствует прямолинейный отрезок (ребро UV графа Г), соединяющий две *неконцевые* вершины U и V нечётной степени, что и доказывает лемму 2.

Удалив из графа Г все прямолинейные рёбра UV, соединяющие неконцевые вершины нечётной степени, приходим к графу Γ', все неконцевые вершины которого имеют чётную степень: $0, 2$ или 4, т. е. оказываемся в условиях леммы 1, которая, как и прежде, влечет за собой теорему 1.

Попутно приведённые доказательства объясняют роль ручки H в задачах 1–3: это аналог начального и конечного звеньев ломаной $A_0A_1 \dots A_n$; наличие ручки H даст возможность выделить среди всех вершин соответствующего графа Г в фазовом пространстве «концевые» вершины F_i спени 1 (подробнее об этом — в разд. 6), а также поможет использовать задачи 1–3 при решении задачи о гвоздях из аннотации.

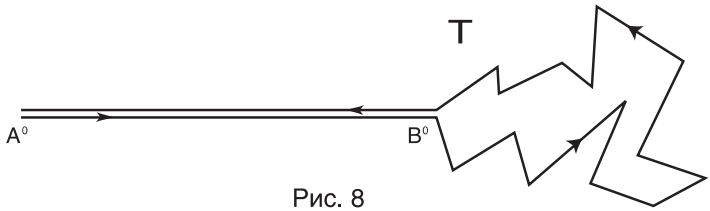


Рис. 8

6. Перемещение отрезка по многоугольнику с ручкой

Чтобы решить задачи 1–3, свяжем с P_H ориентированную ломаную T, состоящую из трёх частей (рис. 8): начальное звено ломаной T — кусок ручки H от точки A^0 до точки B^0, вторая часть T — ориентированный против часовой стрелки контур P многоугольника M с началом и концом в B^0 и, наконец, конечное звено T — кусок ручки H от B^0 до A^0. Точку A^0 выберем *достаточно далеко* от P — так, чтобы её расстояние до P было больше D, где D больше диаметра многоугольника M.

Задача 1, очевидно, имеет решение, если длина d отрезка A_0B_0 больше D.

Поэтому далее предполагается, что $d \in (0; D]$. Кроме того, отрезок A_0B_0 можно без ограничения общности считать примыкающим к A^0, так что далее либо $A_0 = A^0$, либо $B_0 = A^0$.

Построим в фазовом пространстве граф $\Gamma = \Gamma_d$ (d — параметр, $0 < d \leqslant D$); отнесем к Γ_d все фазовые точки (x_1, x_2), $0 \leqslant x_1 < x_2 \leqslant 1$, которые получаются по следующему правилу. Пусть A и B — такие точки ломаной T, что $|AB| = d$, причем при движении из A^0 вдоль T точка A встречается раньше, чем B. Пусть L — длина ломаной T. Длины ломаных A^0A и A^0B (измеренные вдоль T) обозначим через L_1 и L_2 (так что $L_2 - L_1 \geqslant d$) и положим (как и в (3)) $x_i = L_i/L$, $i = 1, 2$. В частности, отрезкам длины d, примыкающим к A^0, соответствуют фазовые точки

$$F_1 = (0, d/L),\ F_2 = (0, 1 - d/L),\ F_3 = (d/L, 1) \text{ и } F_4 = (1 - d/L, 1). \quad (7)$$

Тем самым, в отличие от разд. 5, концевых фазовых точек не две, а четыре (рис. 9).

При $d = D$ и при $d \in (0; D)$, близких к 0, решение задачи 1, очевидно, существует, причем выполняются (соответственно) условия (1) и (2). Иными словами, в первом случае (рис. 9a) Γ_d включает компоненту, содержащую F_1 и F_2, и компоненту, содержащую F_3 и F_4, а во втором случае (рис. 9b) — компоненту, содержащую F_1 и F_4, и компоненту, содержащую F_2 и F_3 (последняя, впрочем, изображает движение отрезка

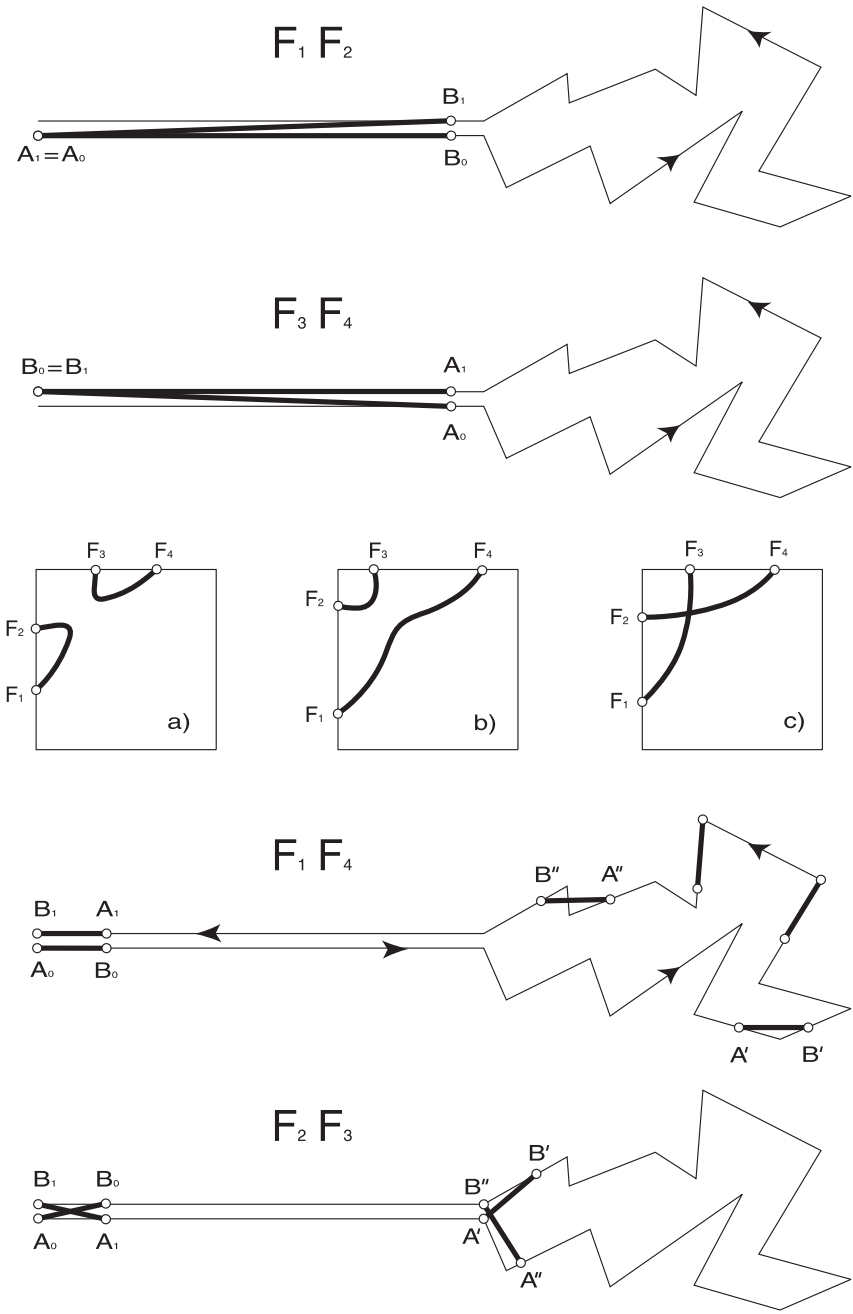

F1 F2
B1
A1 = A0
B0
F3 F4
B0 = B1
A1
A0
F3
F4
F2
F1
a)
b)
c)
F1 F4
B″
A″
B1
A1
A0
B0
A′
B′
F2 F3
B′
B1
B0
B″
A′
A0
A1
A″

Рис. 9

AB, при котором ни A, ни B не обходят P, так что отображения $[0;1]$ в P_H, соответствующие этой компоненте, не являются решением задачи 1).

Решим теперь задачу 1 для любого $d \in (0D]$.

Ясно, что все концевые вершины F_i, $1 \leqslant i \leqslant 4$, графа Γ_d имеют степень 1. Что же касается неконцевых вершин Γ_d, то, почти дословно повторяя доказательство теоремы 1, получаем:

A. *Если ломаная T не содержит ни одной пары параллельных звеньев, расстояние между которыми равно d, то все неконцевые вершины Γ_d имеют четную степень.*

B. *Если ломаная T содержит параллельные звенья, расстояние между которыми равно d, то все неконцевые вершины Γ_d нечётной степени можно разбить на пары вершин (U, V), соединённых в Γ_d прямолинейными отрезками — рёбрами UV. Исключив эти рёбра из Γ_d, приходим к графу Γ'_d, все неконцевые вершины которого имеют чётную степень.*

Выйдя из вершины F_1, будем двигаться по рёбрам графа Γ_d (в случае **A**) или графа Γ'_d (в случае **B**), не проходя ни по одному ребру дважды (пока это возможно); это движение не может закончиться ни в какой вершине чётной степени; тем самым мы дойдем до одной из вершин F_i, $2 \leqslant i \leqslant 4$.

Если это — вершина F_2, то выйдем из F_3 и продолжим двигаться по тем рёбрам, по которым мы ещё ни разу не проходили (по–прежнему не проходя ни по одному ребру дважды, пока это возможно); как и раньше, это движение не может закончиться ни в какой вершине чётной степени; тем самым мы дойдем до единственной оставшейся концевой вершины F_4.

Итак, в рассматриваемом случае граф Γ_d включает компоненту $K_{1,2}$, содержащую точки F_1 и F_2, и компоненту $K_{3,4}$, содержащую F_3 и F_4.

Альтернативная возможность: выйдя из F_1, мы могли дойти до F_4, в этом случае Γ_d включает компоненту $K_{1,4}$, содержащую F_1 и F_4, а также компоненту $K_{2,3}$, содержащую F_2 и F_3.

Наконец, выйдя из F_1, мы могли дойти до F_3 (по пути $K_{1,3}$); но тогда существует и путь $K_{2,4}$ из F_2 в F_4, а поскольку путь $K_{1,3}$ разделяет F_2 и F_4, то пути $K_{1,3}$ и $K_{2,4}$ пересекаются; значит, в этом случае все четыре вершины: F_1, F_2, F_3 и F_4 — принадлежат одной компоненте графа Γ_d (рис. 9c).

Таким образом, из вершины F_1 всегда можно дойти либо до F_2, либо до F_4, а иногда и до каждой из трёх вершины: F_2, F_3 и F_4. Задача 1 решена.

Более того, всё подготовлено и для решения задач 2 и 3.

Определение. Используя обозначение $K_{i,j}$ для связного подграфа графа Γ_d, содержащего F_i и F_j, выделим два класса графов Γ_d:

любой граф первого класса включает непересекающиеся компоненты $K_{1,2}$ и $K_{3,4}$,

любой граф второго класса включает непересекающиеся компоненты $K_{1,4}$ и $K_{2,3}$.

В этих терминах задачи 2 и 3 можно переформулировать следующим образом:

2. *Существует такое w, что при $d > w$ граф Γ_d принадлежит первому классу, а при $d < w$ — второму классу.*

3. *При $d = w$ граф Γ_d не принадлежит ни первому, ни второму классу: все четыре вершины F_i, $1 \leqslant i \leqslant 4$, принадлежат одной компоненте Γ_w.*

Решение задач 2 и 3. Ясно, что при $d \neq e$ графы Γ_d и Γ_e не пересекаются. При этом для концевых вершин Γ_e (обозначим их E_i, $1 \leqslant i \leqslant 4$, чтобы отличить от вершин F_i графа Γ_d) выполняются соотношения, аналогичные (7):

$$E_1 = (0, e/L),\ E_2 = (0, 1 - e/L),\ E_3 = (e/L, 1) \text{ и } E_4 = (1 - e/L, 1),$$

т. е. (рис. 10) при $e < d$ точки E_i расположены ближе, чем F_i, к углам квадрата

$$K = \{(x_1, x_2) \,|\, 0 \leqslant x_i \leqslant 1,\ i = 1, 2\}.$$

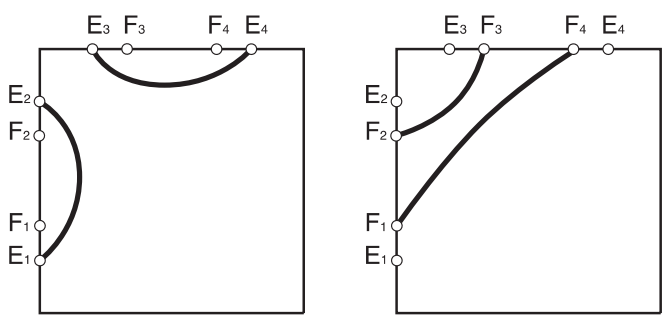


Рис. 10

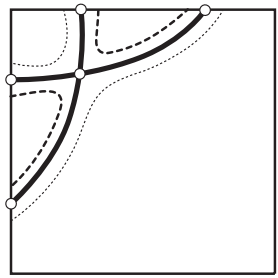


Рис. 11

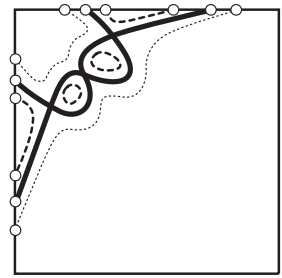


Рис. 12

Поэтому верна

Лемма 3. *Если граф* Γ_e *— первого класса и* $d > e$, *то граф* Γ_d *— тоже первого класса. Если* Γ_d *— второго класса и* $e < d$, *то* Γ_e *— тоже второго класса.*

Доказательство. Иначе Γ_d и Γ_e пересеклись бы, рис. 10.

Следствие. *Можно принять за* w *нижнюю грань тех* d, *для которых граф* Γ_d *принадлежит первому классу или (что то же самое) верхнюю грань тех* d, *для которых* Γ_d *— второго класса. Для этого* w *утверждения задач* 2 *и* 3, *очевидно, выполняются.*

Замечание. Точка $d = w$ является точкой перехода от графов второго класса к графам первого класса; компоненты $K_{i,j}$ для графов Γ_d при d, близких к w (рис. 11), напоминают семейство гипербол $xy = c$ для c вблизи 0. При $d = w$ граф $\Gamma_d = \Gamma_w$ имеет вершину X степени 4, из которой можно попасть по рёбрам Γ_w в любую из вершин F_i, $1 \leqslant i \leqslant 4$; связный подграф Γ_w, содержащий X и F_i, $1 \leqslant i \leqslant 4$, — аналог *сепаратрисы* $xy = 0$, разделяющей гирперболы $xy = c$ для c разных знаков (хотя иногда этот подграф устроен сложнее, например, так, как в случае, когда M — правильный треугольник, рис. 12).

В заключение этого параграфа сформулируем и докажем ещё одну лемму (сравним её с задачами 1, 2), которая понадобится в разд. 7.

Лемма 4. *Пусть существуют такие непрерывные отображения*

$$a:\ t \in [0;1] \to A = a(t) \in P_H, \quad b:\ t \in [0;1] \to B = b(t) \in P_H, \qquad (8)$$

что при $t = 0$ *и* $t = 1$ *отрезок* AB *занимает одно и то же положение* $A_0B_0 \subset H$: $A_0 = a(0) = a(1)$, $B_0 = b(0) = b(1)$, *при изменении* t *от* 0

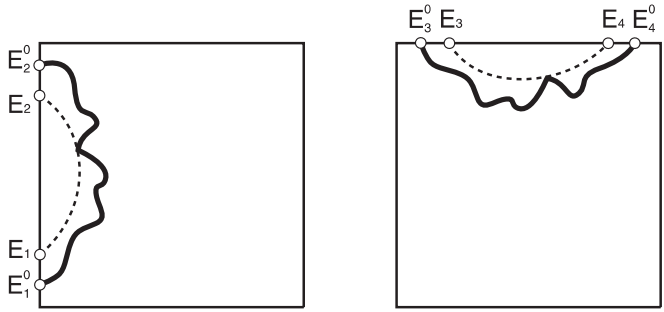


Рис. 13

до 1 *одна из точек* A, B *обходит контур* P *и расстояние между* A *и* B *не обязательно постоянно, но при всех* $t \in [0;1]$ *не превосходит* e (*в частности,* $|A_0B_0| = e_0 \leqslant e$). *Тогда* $e \geqslant w$, *так что существуют отображения, обладающие всеми свойствами отображений* (8) *и, кроме того, такие, что* $|AB| \equiv e$.

Доказательство. В фазовом пространстве перемещение (8) отрезка AB изображается кривой $Q = Q_{1,2}$, соединяющей E_1^0 и E_2^0, или кривой $Q = Q_{3,4}$, соединяющей E_3^0 и E_4^0, где E_i^0 — вершины графа Γ_{e_0} (рис. 13). При $d > e$ эта кривая не пересекается с Γ_d (так как $|AB| \leqslant e < d$). Поэтому $e \geqslant w$: если предположить, что $e < w$, то при $d \in (e; w)$ граф Γ_d — второго класса и, значит, вопреки сказанному, пересекается с кривой Q (либо $K_{1,4}$ пересекает $Q_{1,2}$, либо $K_{2,3}$ пересекает $Q_{3,4}$).

7. Можно ли протащить многоугольник между двумя гвоздями?

Опираясь на лемму 4, решим задачу о гвоздях из аннотации и задачу 4 из разд. 3.

Пусть гвозди G_1 и G_2 вбиты в стол на расстоянии e, и пусть известно, что многоугольник M ширины w_M можно протащить между ними (так что $e \geqslant w_M$).

Когда мы протаскиваем многоугольник между гвоздями, то гвозди неподвижны, а многоугольник M движется. Однако, если жестко связать с M систему координат, то в ней, наоборот, *гвозди* G_1 *и* G_2 *движутся относительно неподвижного многоугольника*:

$$G_1 = G_1(t) \ \text{ и } \ G_2 = G_2(t), \ t \in [0;1]. \tag{9}$$

Для многоугольников не очень сложной формы движение (9) можно представлять себе следующим образом (рис. 14).

На контуре P многоугольника M выбираются две точки: C_1^* и C_2^*. В них к M приделываются две ручки: h_1 и h_2 (полупрямые $C_1^*C_1$ и $C_2^*C_2$), не пересекающие друг друга и не имеющие с P других общих точек, кроме C_1^* и C_2^*. Объединение M, h_1 и h_2 разбивает $\mathbb{R}^2$ на две части: *верхнюю* и *нижнюю*, и движение (9) происходит так, что $G_1(t)$ лежит в замыкании верхней, а $G_2(t)$ — в замыкании нижней части.

Именно так *протаскивание многоугольника между гвоздями* G_1 *и* G_2 трактуется в [1], и исследуемая задача решена в [1] для многоугольников M, которые можно *именно таким способом* протащить между G_1 и G_2.

Но для более сложных многоугольников (рис. 15) при e, близких к w_M движение (9) *нельзя* представить таким образом, и на помощь приходит конструкция P_H из разд. 3.

Чтобы применить её, прежде всего отметим, что если движение (9) происходит так, как в задаче 1, т.е. $G_1(t) = a(t)$, $G_2(t) = b(t)$, то *при выполнении условия* (1) *оказывается, что многоугольник* M *протащили между гвоздями* G_1 *и* G_2.

Значит, *ширина* $w = w(P_H)$ *многоугольника с ручкой* P_H *не меньше ширины* w_M *многоугольника* M. Противоположное неравенство, а с ним и тождество

$$w \equiv w_M, \tag{10}$$

сразу вытекают из следующего утверждения (которое будет доказано ниже):

Если многоугольник M *можно как-нибудь протащить между гвоздями* G_1, G_2, *то это можно сделать и так, чтобы движение* (9) *происходило, как в задаче* 1 *при условии* (1) *и, тем самым, чтобы концы отрезка* G_1G_2 *двигались по* P_H.

Кроме леммы 4, при проверке этого утверждения будет использована ещё одна лемма (лемма 5 *о главном интервале*), к формулировке и доказательству которой мы и приступаем.

Среди всех $t \in [0;1]$ в (9) имеются такие t, что пересечение отрезка G_1G_2 с M пусто, такие t, что G_1G_2 не содержит внутренних точек M, но $G_1G_2 \cap P \neq 0$, и, наконец, на $[0;1]$ можно выделить открытое подмножество Ω таких t, для которых отрезок G_1G_2 содержит *внутренние* точки M.

Лемма 5. *Среди интервалов, составляющих* Ω, *есть такой,* ***главный****, интервал* $(t_0;t_1)$, *что при* $t \in [t_0;t_1]$ *и происходит протаскивание*

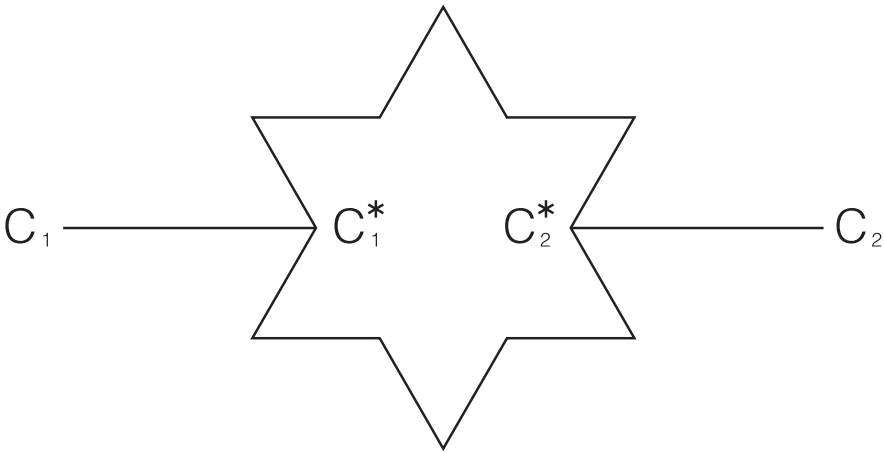


Рис. 14

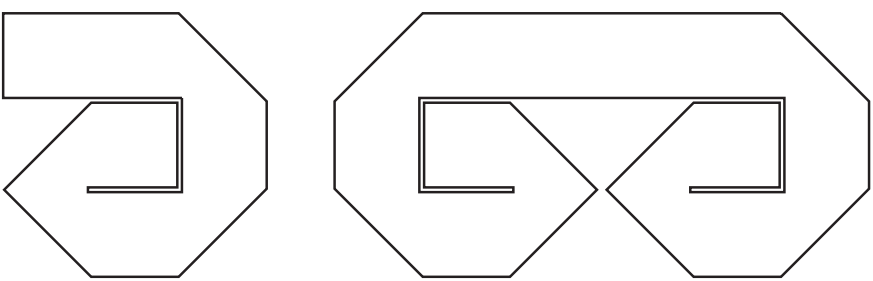


Рис. 15

многоугольника M между G_1 и G_2: для каждой внутренней точки $p \in M$ найдётся $t = t(p) \in (t_0; t_1)$, при котором отрезок G_1G_2 содержит p; при $t \in (t_0; t_1)$ все внутренние точки многоугольника M переходят через отрезок G_1G_2.

Доказательство. Вообще говоря, Ω содержит и другие, *неглавные*, интервалы I. Для t, принадлежащих каждому из них, отрезки G_1G_2 содержат не все, а только часть внутренних точек M; при $t \in I$ эти точки *лишь временно переходят через отрезок G_1G_2, а потом возвращаются обратно*. Поэтому, если бы Ω составляли только такие интервалы I, многоугольник M не был бы протащен между G_1 и G_2. Значит, Ω действительно содержит искомый интервал $(t_0; t_1)$.

Для главного интервала $(t_0; t_1) \subseteq \Omega$ при $t \in [t_0; t_1]$ точки $P \in G_1G_2$, ближайшие к G_1, G_2, обозначим (соответственно) через A и B, $A = A(t)$, $B = B(t)$. Отображения

$$t \in [t_0; t_1] \to A = A(t), \quad t \in [t_0; t_1] \to B = B(t), \tag{11}$$

могут оказаться разрывными (для невыпуклых многоугольников M), но, меняя параметризацию (заменяя «перескоки» $A(t-0)$ в $A(t+0)$ и $B(t-0)$ в $B(t+0)$ «непрерывными переходами»), нетрудно сделать их непрерывными. При каждом из двух граничных значений $t = t_0$ и $t = t_1$ отрезок $AB = A(t)B(t)$ (возможно, вырождающийся в точку — в случае, если $A(t) = B(t)$), очевидно, лежит на границе M, так что существует такая сторона многоугольника s_i, что $A(t_i)B(t_i) \subseteq s_i$, $i = 0, 1$. Поэтому легко доопределить (11) вне $[t_0; t_1]$ и перейти от (11) к непрерывным отображениям (8), удовлетворяющим условиям леммы 4.

Применяя эту лемму, получаем, что существуют отображения $A(t)$ и $B(t)$, обладающие всеми свойствами отображений (8) и, кроме того, такие, что $|AB| \equiv e$.

Полагая $G_1 = A(t)$, $G_2 = B(t)$, получаем, что верна анонсированная выше

Теорема 2. *Если многоугольник M можно как-нибудь протащить между точками G_1, G_2 на расстоянии e, то это можно сделать и так, чтобы движение* (9) *происходило, как в задаче* 1 *при условии* (1)*; концы отрезка G_1G_2 движутся при этом по P_H.*

Так как (по лемме 4) $e \geqslant w$, то при любом $d > e$ соответствующий граф Γ_d принадлежит первому классу и, следовательно, содержит компоненты $K_{1,2}$ и $K_{3,4}$. Отсюда для любого многоугольника M следует

Теорема 3. *Если многоугольник M можно протащить между двумя точками на расстоянии e, то его можно протащить и между точками на расстоянии $d > e$.*

Итак, задача из аннотации решена. Решена и задача 4: как уже отмечено, из теоремы 2 вытекает тождество (10), а поскольку в приведённых выше рассуждениях ручку H можно всюду заменить ручкой H^*, то $w = w^*$ — ширина w многоугольника с ручкой P_H не зависит от расположения ручки H.

8. Доказательство теоремы двойственности

Остается доказать теорему двойственности. Подробно она сформулирована в разд. 1, а её краткая формулировка — следующая:

Теорема 4. $w_M = d_M$.

Доказательство. Вследствие (10) ширина w_M равна d_{max} — максимальному значению длины d отрезка AB, который можно так переместить вдоль ломаной T из положения A_0B_0 в положение $A_1B_1 = B_0A_0$, чтобы выполнялось условие (2); фазовая точка движется при этом по графу Γ_w из F_1 в F_4 (рис. 9).

Граф Γ_w содержит также путь из F_2 в F_3.

Сопоставление перемещений отрезка AB вдоль T, изображаемых путями F_1F_4 и F_2F_3 (нижний фрагмент рисунка 9) делает очевидным равенство $d_{max} = d_M$.

Итак, $w_M = d_M$. Теорема двойственности доказана.

В заключение хочу высказать признательность друзьям и коллегам за помощь при подготовке этой статьи: А. М. Олевский обратил мое внимание на работу [1]; у меня не было этой работы под рукой, и И. В. Арнольд по моей просьбе сканировал и прислал мне её по e-mail; Н. Б. Васильев проявил завидное долготерпение, выслушивая по телефону мои доказательства и обсуждая их со мной; Д. Е. Долгов сделал на компьютере рисунки к статье. Всех их я искренне благодарю.

Список литературы

[1] *Goodman J.E., Pach J. and Yap C.K.* Mountain Climbing, Ladder Moving and the Ring-Width of a Poligon. // Amer. Math. Monthly, Vol.96. 1989. P. 494–510.

[2] *Арнольд В.И.* Обыкновенные дифференциальные уравнения. М.: Наука, 1971.

Проблемы математического образования

Несмотря на долгую традицию преподавания математики в математических кружках и математических классах, вопрос о том, чему следует учить школьников, серьёзно интересующихся математикой, представляется по-прежнему важным и не до конца ясным.

Мы начинаем обсуждение этого вопроса с представления программы «Матшкольник». Во многих математических классах используется система обучения по тематическим циклам задач. Данная программа фактически определяет, что должно входить в такие циклы обязательно, а что должно рассматриваться как дополнительный материал. При этом ни редакция, ни авторы данной программы не считают, что приводимые в ней разделы являются совершенно обязательными и необходимыми. Скорее это образец изложения целей, которых должен достичь школьник, обучающийся в математическом классе. Однако стоит отметить, что, как нам стало известно, в Высшем Математическом Колледже Независимого Московского Университета (негосударственное высшее учебное заведение, специализирующееся на подготовке профессиональных математиков) принято решение рассматривать программу «Матшкольник» как приблизительный стартовый минимум для обучения в МК НМУ.

Редакция надеется, что публикация данной программы вызовет конструктивную реакцию со стороны преподавателей математических классов и других заинтересованных лиц и на страницах нашего сборника будут представлены и другие традиции преподавания в математических классах.

Программа «Матшкольник»

Предлагаемая вниманию читателей программа, по мнению её авторов (А. Вайнтроб при участии А. Шеня и других), содержит минимум сведений, которые должен знать при окончании школы хороший школьник хорошего математического класса. В неё не входят стандартные сведения из школьной программы (решение уравнений, графики, евклидова геометрия и т. п.). Мы старались ограничиваться минимумом новых понятий (скажем, общее понятие группы отсутствует).

Разумеется, эта программа отражает лишь точку зрения её авторов, и возможны другие варианты (например, можно стараться уменьшить её пересечение с программой первого курса за счёт сведений по анализу). Тем не менее, как нам кажется, эта программа соответствует традиции некоторых математических классов и школ (в первую очередь тех, где обучение разделено на «школьную» и «дополнительную» математику).

Уровень требований по каждой теме задаётся образцами задач. Следует иметь в виду, что среди задач устного экзамена есть сильно превосходящие минимальный уровень требований (они давались индивидуально сильным школьникам, легко справившимся с обязательным минимумом).

1. Арифметика и алгебра

1.1. Кольцо целых чисел. Кольца и поля вычетов

1. Делимость целых чисел.
2. Арифметика остатков.
3. Наименьшее общее кратное и наибольший общий делитель.
4. Взаимно простые числа.
5. Алгоритм Евклида.
6. Решение уравнений вида $ax + by = c$.
7. Основная теорема арифметики и её следствия.
8. Бесконечность множества простых чисел.
9. Теорема Ферма–Эйлера.

Образцы задач

1. Доказать, что если $ad + bc$ делится на $a + c$, то и $ab + cd$ делится на $a + c$.
2. Найти наименьшее 60-значное число, делящееся на 101.
3. Найти НОД $(2^{18} - 1, 2^{32} - 1)$.
4. Существует ли число, дающее при делении на 2, 3, 5, 7, 11 остатки 1, 2, 3, 4, 5 соответственно?
5. При каких целых a уравнение $12x + 20y + 30z = a$ имеет целочисленные решения?
6. Число p простое, не равно 2 и 5. Доказать, что дробь $1/p$ периодична и число знаков периода является делителем $p - 1$.
7. См. задачу 1 части 1 письменного экзамена за 1984 год (стр. 203).
8. Сколько существует чисел, которые могут быть остатками при делении точных квадратов на 101?

1.2. Кольцо многочленов

1. Деление многочленов с остатком.
2. Теорема Безу.
3. Корни многочленов и разложение на множители.
4. Конечность числа корней.
5. Многочлены, совпадающие как функции, имеют равные коэффициенты.
6. Интерполяция: существование и единственность.
7. Квадратный трёхчлен. Формула корней.
8. Рациональные корни многочлена с целыми коэффициентами.
9. НОД и НОК многочленов. Алгоритм Евклида.
10. Однозначность разложения на неприводимые (для $\mathbb{R}[x]$).
11. Производная и кратные корни.
12. Симметрические многочлены.

Образцы задач

1. При делении многочлена $P(x)$ на $x-1$ получается остаток 2, а при делении на $x - 3$ — остаток 1. Найти остаток при делении $P(x)$ на $(x - 1)(x - 3)$.
2. Доказать, что НОД $(x^m - 1, x^n - 1) = x^{\text{НОД}(m,n)} - 1$. (См. также задачу 4 части 1 письменного экзамена за 1984 год на стр. 203.)
3. Известно, что $P(x) < 5(x^2 + 1)^3 + 1000$ при всех x. Что можно сказать о степени P?

4. Доказать, что система

$$\begin{cases} x+y+z+t=a \\ x+2y+3z+4t=b \\ x+4y+9z+16t=c \\ x+8y+27z+64t=d \end{cases}$$

при любых a, b, c, d имеет единственное решение.

5. Нарисовать множество тех пар (p,q), при которых квадратный трёхчлен x^2+px+q имеет 2 положительных корня.

6. Доказать, что если число $\sqrt[n]{a}$ рационально (для натуральных a и n), то оно — целое.

7. Найти многочлен $P(x)$ степени 3, для которого $P(1)=1$, $P(2)=5$, $P(3)=0$, $P(4)+P(5)=8$.

8. Многочлен степени 4 имеет 3 действительных корня. Может ли его производная иметь ровно 1 действительный корень?

9. Найти $\displaystyle\sum_{n=1}^{1000} \frac{1}{n(n+1)(n+2)}$.

10. Известно, что $x+y+z$, $xy+yz+xz$, xyz — целые числа. Можно ли утверждать, что число $x^3+y^3+z^3$ — целое?

11. Доказать, что числа $1+\sqrt{2}+\sqrt{3}$ и $\sqrt[3]{2}+\sqrt{5}$ являются корнями ненулевых многочленов с целыми коэффициентами.

1.3. Поле комплексных чисел

1. Комплексные числа и операции над ними.
2. Возможность и однозначность деления.
3. Сопряжённые числа.
4. Основная теорема алгебры (формулировка). Следствие: всякий многочлен из $\mathbb{R}[x]$ разлагается на линейные и квадратные множители в $\mathbb{R}[x]$.
5. Тригонометрическая форма комплексного числа.
6. Геометрический смысл умножения.

Образцы задач

1. Вычислить $(1+i)^{1001}$.
2. Найти произведение и сумму всех корней степени n из числа a.
3. Какой аргумент может иметь число z, если $|z-i|<0{,}5$?

4. Найти многочлен минимальной степени из $\mathbb{R}[x]$, для которого числа $1+i$, 2 и $3-i$ были бы корнями.
5. Доказать, что множество тех z, для которых $Re(1/z)=1$, есть окружность без точки.
6. Какие значения может принимать произведение всех расстояний от точки $\langle 2,0\rangle$ до вершин правильного семиугольника, вписанного в окружность единичного радиуса с центром в 0?
7. На комплексной плоскости даны точки a и b. Где находятся те z, для которых $(z-a)/(z-b)$ — чисто мнимое число?

1.4. Комбинаторика. Группа перестановок

1. Перестановки.
2. Размещения с повторениями.
3. Размещения.
4. Сочетания.
5. Бином Ньютона.
6. Треугольник Паскаля.
7. Формула включений и исключений.
8. Умножение перестановок.
9. Чётные и нечётные перестановки.
10. Разложение перестановок в произведение циклов.

Образцы задач

1. Сколько пятизначных чисел содержат в своей записи цифру 8 и не содержат 0?
2. Сколько решений в натуральных числах имеет уравнение
$$xyz = 2\cdot 3\cdot 5\cdot 7\cdot 11?$$
3. Сколько существует пятизначных чисел, составленных из цифр 1, 2, 3, 4, 5 (каждая входит по одному разу), в которых цифра 5 не стоит на пятом месте?
4. Сколько решений имеет уравнение $x_1+x_2+\cdots+x_k=l$ в целых неотрицательных числах?
5. Найти сумму всех чисел n-ой строки треугольника Паскаля с чётными номерами.
6. Найти знакопеременную сумму всех чисел n-ой строки треугольника Паскаля с чётными номерами ($C_n^0-C_n^2+C_n^4+\ldots$).

7. Доказать, что число разбиений целого положительного числа n на нечётные целые положительные слагаемые равно числу его разбиений на различные целые положительные слагаемые. (Порядок слагаемых считается несущественным.)

8. Существует ли перестановка 7-элементного множества, имеющая порядок 34?

9. Доказать, что всякая чётная перестановка может быть представлена в виде произведения перестановок вида (i, j, k), при которых $i \mapsto j$, $j \mapsto k$, $k \mapsto i$, а остальные числа остаются на месте.

2. Анализ

2.1. Теория множеств

1. Счётные множества.
2. Произведение и сумма счётных множеств.
3. Если A бесконечно, а B счётно, то объединение A и B равномощно A.
4. Теорема Кантора–Бернштейна (формулировка).
5. Пример несчётного множества.
6. Теорема Кантора (для всякого множества существует множество большей мощности).
7. Равномощность прямой и плоскости.

Образцы задач

1. Какую мощность имеет множество всех непрерывных функций на отрезке?

2. Доказать, что $A \times A$ равномощно A, если A — множество всех бесконечных последовательностей нулей и единиц.

3. Доказать, что круг и квадрат (с внутренностями) равномощны.

4. Доказать, что множество действительных чисел равномощно множеству иррациональных чисел.

2.2. Последовательности и пределы

1. Предел последовательности.
2. Единственность предела.
3. Теорема о двух милиционерах.
4. Предел суммы, разности, произведения и частного.
5. Предел отношения показательной и степенной функций.

Образцы задач

1. Доказать, что если $x_n^2 + y_n^2 \to 0$, то $x_n + y_n \to 0$.
2. Доказать, что если последовательность сходится к положительному числу a, то последовательность квадратных корней из её членов сходится к $\sqrt{a}$.
3. Найти пределы $100^n/n!$, $n^{1/n}$, $n^2/2^n$, $((2^n + 3^n + 4^n)/(5^n + 6^n))^{1/n}$.
4. Доказать, что всякая последовательность имеет монотонную подпоследовательность.
5. Имеет ли последовательность $\sin 1$, $\sin 2$, $\sin 3$, ... предел?

2.3. Свойства действительных чисел

1. Точная верхняя грань.
2. Вложенные отрезки.
3. Предел монотонной ограниченной последовательности.
4. Критерий Коши.
5. Покрытие отрезка интервалами.
6. Существование иррациональных чисел.
7. Несчётность множества действительных чисел.

Образцы задач

1. Если M — бесконечное множество точек отрезка, то существует такая точка x, что любой интервал, содержащий x, содержит бесконечно много точек из множества M.
2. Доказать, что если в множестве отрезков любые два имеют общую точку, то существует точка, принадлежащая всем отрезкам этого множества.
3. Если функция f такова, что для любой точки отрезка существует содержащий её интервал, на котором f ограничена, то f ограничена на всем отрезке.
4. Доказать, что если разность между n-м и k-м членами последовательности не превосходит по модулю $1/n + 1/k$, то эта последовательность сходится. Что можно сказать, если эта разность не превосходит по модулю $1/(nk)$?

2.4. Числовые ряды

1. Сходимость и абсолютная сходимость.
2. Сумма и произведение рядов.
3. Признак сравнения.
4. Интегральный признак.
5. Геометрическая прогрессия.
6. Гармонический ряд.

Образцы задач

1. Доказать, что если ряды с членами x^2 и y^2 сходятся, то и ряд с членами $x_n y_n$ сходится.
2. При каких x сходится ряд с членами $n^2 x^n$?
3. Доказать, что функция $f(x)$, равная сумме ряда $x^n/n!$, n — целое неотрицательное, определена при всех x и $f(x+y) = f(x)f(y)$.
4. При каких p сходится ряд с членами $1/(n^p + n)$?
5. Вычислить сумму ряда $\sum_n 1/n^3$ с точностью до 0,01.

2.5. Непрерывность на прямой

1. Различные определения непрерывной функции на прямой.
2. Достижение максимума.
3. Прохождение нуля.
4. Равномерная непрерывность.
5. Непрерывность элементарных функций (многочлены, тригонометрические функции, логарифм).

Образцы задач

1. Доказать, что если f — непрерывная функция, отображающая отрезок $[0;1]$ в себя, то существует такое x, что $f(x) = x$.
2. Построить функцию на действительной прямой, множество точек разрыва которой состоит из всех точек вида $1/n$ при целых положительных n.
3. Функция f определена при неотрицательных x так: $f(x) = (1 + x^3)/2^x$. Будет ли она равномерно непрерывной?
4. Какие множества могут быть множествами значений непрерывных на интервале $(0;1)$ функций?
5. Доказать непрерывность функции $f(x) = \sum_n n^2 x^n$ на отрезке $[0;0{,}5]$.
6. Первый член последовательности равен 1, а каждый следующий — квадратному корню из суммы предыдущего члена и числа 2. Имеет ли эта последовательность предел?

2.6. Дифференцирование на прямой

1. Определение производной.
2. Производные суммы, произведения, частного.
3. Производная сложной функции.
4. Производные элементарных функций.

5. Теоремы Ролля и Лагранжа.
6. Монотонность и первая производная.
7. Выпуклость и вторая производная.
8. Применение выпуклости к доказательству неравенств.
9. Формула Тейлора.

Образцы задач

1. Известно, что уравнение $f(x) = a$ имеет n решений. Какое число решений может иметь уравнение $f'(x) = 0$?
2. Может ли производная всюду дифференцируемой функции не быть непрерывной?
3. Известно, что $|f(y) - f(x)| \leqslant (y - x)^2$. Доказать, что f — константа.
4. Известно, что $h(x) = f(f(x))$, $f(2) = 2$, производная $h'(2)$ равна 3. Чему может быть равно $f'(2)$?
5. Найти число решений уравнения $x^3 - x + a = 0$ в зависимости от a.
6. Найти касательную к кривой $x^3 + x + y^3 + y = 2$ в точке $\langle 1, 0\rangle$.
7. Доказать, что среднее геометрическое не больше среднего арифметического, пользуясь выпуклостью логарифма.
8. Найти предел отношения $(\sin x - \operatorname{tg} x)/x^3$ при $x \to 0$.

2.7. Интеграл

1. Интеграл непрерывной функции по отрезку.
2. Первообразная непрерывной функции.
3. Теорема Ньютона–Лейбница.
4. Интегрирование по частям и замена переменной.

Образцы задач

1. Найти $g'(1)$ и $g'(2)$, если $g(x) = \int_x^{x^2} (\sin t)/t\, dt$
2. Найти точную верхнюю грань чисел $\int_0^1 xf(x)\, dx$ по всем непрерывным на $[0; 1]$ функциям, для которых $\int_0^1 f(x)\, dx \leqslant 2$.
3. Функция f непрерывна на $[0; 1]$. Доказать, что $\int_0^1 f(x) \sin nx\, dx \to 0$ при $n \to \infty$.
4. Вычислить $\int x \ln x\, dx$.
5. Доказать, что последовательность $1 + 1/2 + \cdots + 1/n - \ln n$ монотонна и ограничена.

3. Геометрия

3.1. Группы преобразований плоскости и их комплексный смысл

1. Движения: перенос, поворот, симметрии.
2. Вычисление композиций различных видов движений.
3. Формулировка теоремы Шаля о классификации движений.
4. Преобразования подобия.
5. Композиции гомотетий и движений.
6. Дробно-линейные преобразования комплексной плоскости.
7. Инверсия.

Образцы задач

1. Если фигура имеет два центра симметрии, то она имеет и третий.

2. Во что переходит треугольник с вершинами $1+i$, $2-3i$, $4-i$ при повороте на 120° вокруг $2-2i$?

3. Найти все движения, перестановочные с поворотом на 90° вокруг данной точки.

4. При каких a, b преобразование $z \mapsto a\bar{z} + b$ является симметрией?

5. При каких a, b преобразование $z \mapsto az + b$ является гомотетией?

6. Две карты одной местности разных масштабов положены друг на друга. Доказать, что их можно проколоть иглой, отметив на обеих картах одну и ту же точку местности.

7. Даны три окружности разных радиусов. К каждой паре проведены внешние касательные и взята точка их пересечения. Доказать, что три этих точки лежат на одной прямой.

8. Доказать, что комплексные числа a, b, c, d лежат на одной прямой или окружности, если $\dfrac{a-c}{a-d} : \dfrac{b-c}{b-d}$ действительно.

9. Найти все дробно-линейные преобразования, отображающие верхнюю полуплоскость на себя.

10. Доказать, что центры правильных треугольников, построенных на сторонах произвольного треугольника, образуют правильный треугольник.

11. Дана точка и две окружности. Построить окружность, проходящую через данную точку и касающуюся данных окружностей.

3.2. Геометрия векторных пространств

1. Координатное пространство и его подпространства.
2. Системы линейных уравнений и их геометрический смысл.
3. Теорема: однородная система, в которой неизвестных больше, чем уравнений, имеет ненулевое решение.
4. Линейная зависимость.
5. Базисы. Размерность.
6. Пересечение и сумма подпространств: соотношение размерностей.
7. Скалярное произведение.
8. Неравенство Коши–Буняковского.
9. Неравенство треугольника.
10. Угол между векторами.

Образцы задач

1. Найти размерность минимального подпространства в пятимерном пространстве, содержащего вектора

$$\langle 1,1,5,1,1\rangle,\ \langle 0,2,4,2,1\rangle,\ \langle 0,3,3,3,1\rangle,\ \langle 0,4,2,4,1\rangle,\ \langle 0,5,1,6,2\rangle.$$

2. Доказать, что векторы $\langle 1,2,\dots,n\rangle, \langle 1^2,2^2,\dots,n^2\rangle, \dots, \langle 1^n,2^n,\dots,n^n\rangle$ линейно независимы.

3. Доказать, что любая последовательность, удовлетворяющая соотношению $a_{n+2}=a_{n+1}+2a_n$, имеет вид $a_n=A2^n+B(-1)^n$ при некоторых A и B.

4. При каких условиях на числа a, b, c, d можно найти вектора u и v в пространстве, для которых $(u,u)=a$, $(u,v)=b$, $(v,u)=c$, $(v,v)=d$?

5. Найти площадь треугольника, заданного координатами его вершин в пространстве.

6. Найти расстояние от точки пространства, заданной её координатами, до плоскости, заданной коэффициентами определяющего её линейного уравнения.

7. Длина каждого из трёх векторов пространства равна 1, а скалярное произведение любой пары векторов равно $-0{,}5$. Доказать, что эти вектора линейно зависимы.

4. Письменный экзамен 1984 года

Письменный экзамен по программе «Матшкольник», проходивший в конце 1983/84 учебного года, состоял из двух частей: по арифметике и алгебре и по анализу. На обе части вместе было предоставлено 6 часов.

Часть 1. Арифметика и алгебра

1. Решить в целых положительных числах уравнение
$$\frac{1}{x+\dfrac{1}{y+\dfrac{1}{z}}}=\frac{7}{27}.$$
2. Существует ли число, записываемое n единицами подряд и делящееся на 49, если $0<n<45$?
3. На плоскости изображены правильный m-угольник и правильный n-угольник. Доказать, что если m и n взаимно просты, то можно построить с помощью циркуля и линейки правильный mn-угольник.
4. Найти наибольший общий делитель многочленов $x^{480}-1$ и $x^{36}+1$.
5. Многочлен $P(x)$ удовлетворяет тождеству $P(x)=P(1-x)$. Доказать, что есть такой многочлен $M(y)$, что $P(x)=M((x-0{,}5)^2)$.
6. Известно, что $z+1/z=2\cos a$. Доказать, что $z^n+1/z^n=2\cos na$.
7. Найти сумму квадратов сторон и диагоналей правильного 7-угольника, вписанного в единичную окружность на комплексной плоскости.
8. Найти размерность минимального подпространства, содержащего строки $\langle 1,a,-1,2\rangle$, $\langle 2,-1,a,5\rangle$, $\langle 1,10,-6,1\rangle$ (a — параметр).
9. Матрицу 2×2 будем записывать строчкой из четырёх входящих в неё чисел. Доказать, что отображение, переводящее матрицу X в матрицу AX является линейным оператором и найти его матрицу (размера 4×4). Здесь A — некоторая фиксированная матрица размера 2×2.
[Примечание. В вариант программы, по которой проводился экзамен, входило понятие матрицы и линейного оператора.]
10. Какова вероятность угадать ровно 4 номера в игре «Спортлото 6 из 49»? [В этой игре надо выбрать 6 чисел среди $1,2,\ldots,49$, во время розыгрыша выбираются другие шесть чисел и считается количество общих чисел в этих двух шестёрках.]
11. Сколько решений имеет уравнение $x_1+x_2+\cdots+x_{10}=100$, где переменные принимают натуральные значения, большие 2?
12. Доказать, что нельзя, поворачивая грани кубика Рубика, добиться того, чтобы
(а) один рёберный (на середине ребра) кубик перевернулся, а остальные остались в прежнем положении и по-старому ориентированными;

(б) угловые кубики одной из граней циклически переставились, а остальные остались на своих местах (ориентация кубиков может поменяться).

Часть 2. Математический анализ

1. Найти предел последовательности, заданной соотношениями $a_1 = 1$, $a_{n+1} = 0{,}5(a_n + 2/a_n)$.
2. Доказать, что можно выбрать знаки так, чтобы равенство $1 \pm \frac{1}{2} \pm \pm \frac{1}{3} \pm \cdots = 5$ стало верным.
3. Является ли равномерно непрерывной на действительной оси функция (а) e^{-x}; (б) $\dfrac{1}{1+x^2}$?
4. Доказать, что $\sin x > x - x^3/6$ при $0 < x < \pi/2$.
5. Доказать, что множество точек разрыва монотонной функции не более, чем счётно.
6. $P(x)$ — многочлен, разлагающийся на линейные множители с действительными коэффициентами, причем $P'(0) = P''(0) = 0$. Доказать, что $P(0) = 0$.
7. Сходится ли ряд $\sum 1/(n \ln n)$?
8. Доказать, что предел интегралов по любому отрезку от функций $e^{x^2} \cos nx$ при n, стремящемся к бесконечности, равен 0.
9. Функция определена на всей числовой оси, причем у каждой точки есть окрестность, в которой функция монотонно возрастает. Доказать, что функция монотонно возрастает на всей числовой оси.
10. Существует ли функция, непрерывная во всех иррациональных точках и разрывная во всех рациональных?

5. Письменный экзамен, май 1985

Часть 1. Арифметика и алгебра

1. (а) Числа p и q простые. Доказать, что если $2^p - 1$ делится на q, то $q - 1$ делится на $2p$; (б) просто ли число $2^{13} - 1$?
2. Число x называется квадратичным вычетом по простому модулю p, если оно сравнимо с некоторым точным квадратом по этому модулю. Доказать, что произведение вычета, не делящегося на p, и невычета является невычетом по модулю p.
3. Решить уравнение $(x + iy)^5 = x - iy$.
4. (а) Найти сумму $\sin(\pi k/2n)$ по всем $k = 1, \ldots, 2n$; (б) Найти произведение $\sin(\pi k/2n)$ по всем $k = 1, \ldots, n$.

5. Найти сумму коэффициентов многочлена $(x^2 - x + 1)^{100}$.
6. При каком n многочлен $a^n + a^{n-1}b + \ldots + ab^{n-1} + b^n$ делится на $a^2 + ab + b^2$ (как многочлен над $\mathbb{C}$)?
7. Найти максимальный порядок перестановки 10-элементного множества.

Часть 2. Анализ

1. Найти предел суммы $1/n + 1/(n+1) + \cdots + 1/2n$ при $n \to \infty$.
2. Найти точную верхнюю грань всех чисел вида $\sin x + \cos\sqrt{x}$ по всем действительным x.
3. Что больше: π^e или e^π?
4. Каждый следующий член последовательности равен синусу предыдущего. Найти её предел.
5. Найти первообразную функции $1/\sin x$.
6. Дана ограниченная выпуклая фигура площади S. Доказать, что можно провести две перпендикулярные прямые, делящие её на части площади $S/4$.
7. Дважды дифференцируемая на положительной полуоси функция при $x \to \infty$ стремится к 0 и имеет ограниченную вторую производную. Доказать, что её первая производная стремится к 0.
8. Доказать, что множество значений действительной функции действительного аргумента в точках (нестрогих) максимумов не более, чем счётно.
9. Сходится ли ряд $\sum n(\sin n)/2^n$?

Часть 3. Геометрия

1. Был пятиугольник. На его сторонах построили вовне правильные треугольники и отметили их центры, а всё остальное стерли. Восстановить пятиугольник.
2. Найти все преобразования подобия, перестановочные с осевой симметрией.
3. Одна окружность лежит внутри другой. Доказать, что существует инверсия, переводящая окружности в концентрические.
4. При каких комплексных a, b, c, d преобразование $z \mapsto (a\bar{z}+b)/(c\bar{z}+d)$ — инверсия?
5. Сумма трёх углов равна 2π. Доказать, что сумма их косинусов не меньше $-3/2$.

6. В четырёхмерном пространстве рассмотрим линейную оболочку векторов $\langle 1,2,0,1\rangle$ и $\langle 1,1,1,0\rangle$, а также линейную оболочку векторов $\langle 1,0,1,0\rangle$ и $\langle 1,3,0,1\rangle$. Найти размерности суммы и пересечения этих подпространств.

6. Задачи устных экзаменов

6.1. Алгебра

1. (а) Найти максимальный порядок чётной перестановки множества из 12 элементов. (б) Сколько существует чётных перестановок этого множества? (в) То же для множеств из 30 и 10 элементов.

2. Найти коэффициент при x^{179} в $(1-x+x^4)^{60}(1+x+x^4)^{60}$.

3. Слонопотам ходит по бесконечной шахматной доске, как конь, только не на $\langle 1,2\rangle$, а на $\langle m,n\rangle$. При каких m и n слонопотам может попасть из начальной клетки в соседнюю?

4. (а) Уравнения $x^2-a=yp$ и $x^2-b=yp$ (a,b,p фиксированы, p — простое число) неразрешимы в целых числах. Доказать, что уравнение $x^2-ab=yp$ разрешимо в целых числах. (б) Сколько вычетов по модулю p являются квадратами ?

5. (а) Доказать, что многочлен, являющийся неприводимым в кольце многочленов с рациональными коэффициентами, не имеет кратных корней.
(б) Доказать неприводимость над $\mathbb{Q}$ многочлена $x^6+x^5+\ldots+1$.

6. Найти $\sum_{n=1}^{100}(\cos nx)/2^n$.

7. (а) Доказать, что многочлен $P(x)=(x-a_1)(x-a_2)\ldots(x-a_6)-1$ не разлагается на множители с рациональными коэффициентами.
(б) Тот же вопрос для многочлена $(x-a_1)^{\alpha_1}(x-a_2)^{\alpha_2}\ldots(x-a_n)^{\alpha_n}-1$, где α_i взаимно просты.

8. Многочлен с действительными коэффициентами принимает в целых точках целые значения. Доказать, что он представляется в виде суммы многочленов $q_i(x)=x(x-1)(x-2)\ldots(x-i+1)/i!$ с целыми коэффициентами.

9. Найти $\sum_{i=1}^{p-1} i^k \bmod p$ для произвольного k и простого p.

10. (а) Каждая из граней куба раскрашена в один из четырёх цветов. Сколько существует различных раскрасок? (Раскраски, отличающиеся поворотом куба, считаются одинаковыми.) (б) Тот же вопрос для 6 цветов. (в) Тот же вопрос для 3 цветов.

11. (а) Найти 1985^{1000} по модулю 77. (б) Найти две последние цифры числа 77^{1000}.

12. (а) Доказать, что корни производной комплексного многочлена лежат в любой выпуклой фигуре, содержащей корни исходного многочлена.
(б) Доказать, что если $a_1, \ldots, a_n$ и z — комплексные числа, причём $\sum_i 1/(z-a_i) = 0$, то z лежит в любой выпуклой фигуре, содержащей все a_i.

13. Доказать, что для любых двух многочленов $P(t)$ и $Q(t)$ существует ненулевой многочлен от двух переменных $R(x,y)$, для которого $R(P(t),Q(t)) = 0$ при всех t.

14. Пусть $z + 1/z = 2\cos\varphi$. Найти $z^n + 1/z^n$. (См. также задачу 6 части 1 письменного экзамена за 1984 год.)

15. Найти остаток от деления $x^{243} + x^{81} + x^{27} + x^9 + x^3 + x$ на $x^2 - 1$.

16. Можно ли за нечётное число перестановок двух соседних чисел получить из последовательности $1, 2, 3, \ldots, 101$ последовательность $101, 100, 99, \ldots, 2, 1$?

17. (а) Сколько существует чисел, взаимно простых с 2400 и меньших 2400? (Обозначение: $\varphi(2400)$.) (б) То же для числа $p_1^{\alpha_1} \cdot \ldots \cdot p_n^{\alpha_n}$ (p_i — простые). (в) Доказать, что $\varphi(ab) = \varphi(a)\varphi(b)$ для взаимно простых a и b.

18. (а) Доказать, что многочлен, инвариантный относительно чётных перестановок переменных, является суммой симметрического и кососимметрического. (б) Найти размерность пространства кососимметрических многочленов степени 5 от 3 переменных. (в) Найти размерность пространства многочленов от 10 переменных, все одночлены которых имеют суммарную степень 4. (г) Доказать основную теорему о симметрических многочленах. (д) Доказать, что если $x_1, \ldots, x_n$ — список корней многочлена P степени n, то произведение $\prod_{i \neq j}(x_i - x_j)$ полиномиально выражается через коэффициенты многочлена P.

19. (а) Найти длину периода в десятичной записи $1/41$. (б) См. задачу 1.1.6.

20. (а) Доказать, что значение комплексного многочлена степени меньше n в центре правильного n-угольника равно среднему арифметическому его значений в вершинах. (б) Доказать, что если корни многочлена степени n расположены в вершинах правильного n-угольника, то его производная в центре равна 0. Найти этот многочлен. (в) Доказать единственность разложения дроби $1/P(x)$, где $P(x)$ имеет лишь действительные корни, на простейшие дроби.

21. Сколько решений имеет сравнение $x^{11} \equiv 23 \pmod{23}$ (среди чисел от 0 до 22)?

22. Доказать, что $(p-1)! + 1$ делится на p при простом p.

23. (а) Выразить $\cos 77x$ через $\cos x$. (б) Найти многочлен с целыми коэффициентами, у которого числа $\sin(\pi/99), \sin(2\pi/99), \ldots, \sin(98\pi/99)$ являются корнями.

24. (а) $1 + 1/2 + 1/3 + \ldots + 1/(p-1) = m/n$, где p простое. Доказать, что m делится на p. (б) Доказать, что частные суммы гармонического ряда не являются целыми числами.

25. (а) Если сравнение $x^2 + 1 \equiv 0$ разрешимо по простому модулю p, то p даёт остаток 1 при делении на 4. (б) Доказать, что верно и обратное.
(в) Доказать, что сравнение $x^2 - 1 \equiv 0 \pmod{p}$ всегда имеет ровно 2 решения (среди чисел $0, \ldots, p-1$).

26. Можно ли, вращая кубик Рубика, повернуть угловой кубик, оставив остальные в исходном положении?

27. (а) Число p — простое. Доказать, что найдутся три числа x, y, z, не все делящиеся на p, сумма квадратов которых делится на p. (б) Многочлен $P(x, y, z)$ — сумма одночленов суммарной степени 2 с целыми коэффициентами. Доказать, что существуют x, y, z, не все делящиеся на p, при которых $P(x, y, z)$ делится на p. (Число p — простое.)

28. Можно ли вернуть в исходное положение фишки игры в 15, предварительно переставив 14 и 15?

29. (а) Доказать, что если расширение поля $\mathbb{Q}$ конечномерно как векторное пространство над $\mathbb{Q}$, то все элементы расширения — алгебраические.
(б) Доказать конечномерность $\mathbb{Q}(\sqrt[3]{2}, \sqrt{5})$. (в) Доказать, что сумма, произведение и частное алгебраических чисел — алгебраические.

30. Разложить на множители $x^{p-1} - 1$ в поле вычетов по простому модулю p.

31. Найти сумму $\sum_i |MA_i|^2$, если M — точка окружности единичного радиуса, а A_i — вершины вписанного в неё правильного n-угольника.

32. Числа $p_1, \ldots, p_n$ — различные простые, $a_1, \ldots, a_n$ — произвольные целые. Доказать, что существует целое число a, сравнимое с a_i по модулю p_i для любого i.

33. Доказать, что система

$$\begin{cases} x_1 + x_2 + \ldots + x_{10} = a_1 \\ x_1 + 2x_2 + \ldots + 10x_{10} = a_2 \\ \qquad\qquad\qquad\qquad \ldots \\ x_1 + 2^9 x_2 + \ldots + 10^9 x_{10} = a_{10} \end{cases}$$

разрешима при любых $a_1, a_2, \ldots, a_{10}$. (См. также задачу 1.2.4.)

34. Многочлены $P(x, y)$ и $Q(x, y)$ таковы, что $P(ab, a + b) = Q(ab, a + b)$ для всех a и b. Доказать, что $P = Q$.

35. Найти (а) $\sum_{k=0}^{n} (C_n^k)^2$; (б) $\sum_{k=0}^{n} kC_n^k$.

36. Семь белых или чёрных бусинок нанизывают на окружность. Сколько различных ожерелий можно получить ?

37. Существует ли многочлен P с целыми коэффициентами, для которого $P(5) = 13$, $P(10) = 8$, $P(13) = 21$?

38. Разложить на неприводимые целочисленные множители многочлены $x^{12} - 1$, $x^4 + 4$.

39. При каких действительных p и q многочлен $x^2 + px + q$ имеет ровно два различных вещественных корня?

40. Найти максимум $|z|$ при $|(z+1)/z| = a$.

41. Функция $P(x, y)$ двух вещественных переменных с вещественными значениями — многочлен от x при любом фиксированном y и многочлен от y при любом фиксированном x. Доказать, что P — многочлен.

42. (а) Доказать, что все коэффициенты многочлена $(x-1)(x-2)\cdot\ldots\cdot(x-99)(x-100)$ делятся на 101 (кроме старшего коэффициента, равного 1, и свободного члена). (б) Найти остаток от деления свободного члена на 101.

43. Многочлен с действительными коэффициентами неотрицателен при всех действительных аргументах. Доказать, что его можно представить в виде суммы квадратов двух многочленов с действительными коэффициентами.

44. Доказать, что группа вращений куба изоморфна группе перестановок множества из 4 элементов.

45. Найти сумму $\varphi(d)$ по всем делителям d данного натурального n. (Здесь $\varphi(k)$ — количество чисел от 1 до k, взаимно простых с k.)

46. Число p — простое, отличное от 2 и 5. Доказать, что существует делящееся на p число, десятичная запись которого состоит из одних единиц.

47. Доказать, что если нечётное натуральное число единственным образом разлагается в сумму двух квадратов, то оно простое.

48. Многочлен с действительными коэффициентами принимает целые значения в целых точках. Могут ли все эти значения быть простыми?

6.2. Анализ

1. Первый член последовательности a_n равен 1, а каждый следующий — синусу предыдущего. Найти такое λ, что предел $\lim n^\lambda a_n$ существует и не равен 0.

2. (а) Пусть M — максимум модуля производной функции f на отрезке $[0; 2\pi]$. Доказать, что $\left|\int_0^{2\pi} f(x)\cos nx\,dx\right| < 2\pi M/n$.

3. (а) Функция f непрерывна на прямой. Известно, что при любом x предел последовательности $a_n = f(nx)$ равен 0. Можно ли утверждать, что $f(x) \to 0$ при $x \to \infty$? (б) Тот же вопрос для $f(n+x)$ вместо $f(nx)$.

4. Сходится ли ряд $\sum_n (\cos n)/n$?

5. (а) Найти предел последовательности $a_n = 1/(1+1/(1+\ldots(1+1/1)\ldots))$ (n дробей). (б) Доказать существование предела последовательности $a_n = 1/(x_1+1/(x_2+\ldots(x_{n-1}+1/x_n)\ldots))$ (x_i — любая последовательность положительных целых чисел).

6. Известно, что $f(x) = o(x^n)$ при $x \to 0$ (т. е. предел $f(x)/x^n$ при $x \to 0$ равен 0). Следует ли отсюда, что $f^{(n)}(0)$ существует и равно 0? (б) Тот же вопрос, если известно, что все производные в точке 0, вплоть до $(n-1)$-ой, существуют. (в) Тот же вопрос, если эти производные непрерывны.

7. Функции f_n непрерывны на $[0;1]$ и для любого x из $[0;1]$ предел $f_n(x)$ равен $f(x)$. Может ли функция f быть (а) разрывной хотя бы в одной точке? (б) разрывной всюду? (в) разрывной хотя бы в одной точке, если сходимость равномерна?

8. Функция $f : \mathbb{R} \to \mathbb{R}$ бесконечно дифференцируема, и в каждой точке хотя бы одна из производных равна 0. (а) Доказать, что существует отрезок, на котором f — многочлен. (б) Доказать, что f — многочлен.

9. Доказать, что $x > \sin x > 2x/\pi$ при $0 < x < \pi/2$.

10. Ряд из действительных чисел сходится. Может ли ряд из их кубов расходиться?

11. Функция f на отрезке $[a;b]$ называется хорошей, если множество сумм вида $\sum |f(x_{i+1}) - f(x_i)|$ (x_i — возрастающая конечная последовательность точек отрезка $[a;b]$) ограничено. (а) Всякая ли дифференцируемая функция — хорошая? (б) Доказать, что дифференцируемая функция с ограниченной производной — хорошая. (в) Доказать, что хорошими являются те и только те функции, которые могут быть представлены в виде разности двух неубывающих функций.

12. (а) Бывает ли на прямой не равная 0 бесконечно дифференцируемая функция, равная 0 вне некоторого отрезка? (б) Существует ли бесконечно дифференцируемая функция f на прямой, не равная тождественно 1, для которой $f(f(x)) = 1$ при всех x?

13. (а) Доказать, что если $\lim a_n = a$, то $s_n = (a_1 + \ldots + a_n)/n$ также сходится к a. (б) Верно ли обратное?

14. (а) Может ли $\mathbb{Q}$ быть множеством точек непрерывности некоторой функции f? (б) Доказать, что любое замкнутое множество является множеством точек разрыва некоторой функции.

15. Может ли $\mathbb{R} \setminus \mathbb{Q}$ быть объединением счётного числа замкнутых множеств?

16. (а) Доказать, что всякую функцию, непрерывную на замкнутом подмножестве вещественной прямой, можно продолжить до всюду непрерывной функции. (б) Множества X и Y — замкнутые непересекающиеся подмножества $\mathbb{R}$. Доказать, что существует дифференцируемая на прямой функция, для которой X и Y являются прообразами 0 и 1.

17. Рассмотрим множество M всех дважды дифференцируемых функций f, для которых $f(0) = f(1) = 0$ и $|f''(x)| \leqslant C$ для любого x. Найти $\sup\{|f(x)|\}$ по всем $x \in \{0, 1\}$ и $f \in M$.

18. (а) Доказать, что $n! > (n/3)^n$. (б) Доказать, что $\ln(n!)/\ln((n/e)^n) \to 1$ при $n \to \infty$.

19. Доказать, что функция, имеющая счётное число точек разрыва на отрезке, интегрируема по Риману.

20. Доказать, что для любых $a_1, \ldots, a_n$ и $b_1, \ldots, b_n$ уравнение $\sum_{k=1}^{n}(a_k \cos kx + b_k \sin kx) = 0$ имеет решение.

21. Может ли ряд Тейлора функции сходиться в некоторой точке к числу, отличному от значения функции в этой точке?

22. Доказать, что производная дифференцируемой на интервале функции вместе с любыми двумя значениями принимает и все промежуточные.

23. Пусть $g(x) = x + \sin x$. Найти предел $g(g(g \ldots g(x) \ldots))$ (n раз) при $n \to \infty$.

24. Всякое ли замкнутое множество мощности континуум содержит интервал?

25. Функция $f : \mathbb{R} \to \mathbb{R}$ непрерывна и $f(f(x)) = x$ при всех x. Доказать, что существует такое x, что $f(x) = x$.

26. Доказать, что любая последовательность содержит монотонную подпоследовательность.

27. При каких x сходится ряд $\sum_{n=1}^{\infty} n^3 x^n$?

28. (а) Доказать, что равномерно непрерывная на интервале функция ограничена. (б) Верно ли обратное?

29. Функция f дифференцируема на $(0; 1)$ и $f' = f$. Найти все такие f.

30. (а) Найти все непрерывные $f : \mathbb{R} \to \mathbb{R}$, для которых $f(a + b) = f(a) + f(b)$ и $f(ab) = f(a)f(b)$. (б) Та же задача без требования непрерывности.

31. Для каких x последовательность $\sin nx$ сходится? (См. также задачу 2.2.5.)

32. Последовательность задана формулами $a_1 = 1$ и $a_{n+1} = (a_n + 2/a_n)/2$. Доказать, что предел последовательности равен $\sqrt{2}$. Оценить скорость сходимости. (См. также задачу 1 части 2 письменного экзамена за 1984 год на стр. 204.)

33. Построить взаимнооднозначное соответствие между множеством $\mathbb{Q}$ рациональных чисел и множеством чисел вида $m/2^n$ для целых m и n, сохраняющее порядок.

34. Построить функцию f, являющуюся отношением двух многочленов с рациональными коэффициентами, для которой последовательность $x_1 = 1$, $x_{n+1} = f(x_n)$ сходится к $\sqrt[3]{2}$.

35. Известно, что $x_{n+2} = x_n + x_{n+1}$, $x_1 = 5$, $x_2 = 7$. Найти x_{1001}/x_{1000} с точностью до 0,001.

36. При каких α, β, γ точки $\langle\{\alpha n\}, \{\beta n\}, \{\gamma n\}\rangle$ (при $n = 1, 2, 3, \ldots$) плотны в единичном кубе? ($\{s\}$ — дробная часть s.)

37. Найти все непрерывные на $(0; +\infty)$ функции f, для которых $f(2x) = 2f(x)$, $f(3x) = 3f(x)$ при всех положительных x.

38. Найти точную верхнюю грань множества значений функции $\sin x + \sin \sqrt{2}x$.

39. Периодична ли функция $\sin x + \sin \sqrt{2}x$?

40. Доказать, что множество точек разрыва первого рода (когда существуют односторонние пределы) любой функции не более, чем счётно.

41. Найти функцию, принимающую все рациональные значения на любом интервале с рациональными концами.

42. Доказать, что выпуклая функция (у которой любой кусок графика лежит под стягивающей его хордой) непрерывна.

43. Функция $f : \mathbb{R} \to \mathbb{R}$ дважды дифференцируема всюду, положительна на интервале $(a; b)$; $f(a) = f(b) = 0$, $f(x) + f''(x) < 0$ при $a < x < b$. Доказать, что $|b - a| \leqslant \pi$.

44. Найти производную функции $f(x) = x^x$.

45. Доказать, что функция $f(x) = x - x^2/2 + x^3/3 - x^4/4 + \ldots$ определена и дифференцируема на интервале $(-1; 1)$. Найти f'.

46. Найти бесконечно дифференцируемые функции f, g на прямой, для которых множество всех точек $\langle f(t), g(t)\rangle$ есть граница треугольника.

47. Положительная непрерывно дифференцируемая на прямой функция f такова, что $|f'(x)| \leqslant |f(x)|$. Доказать, что $|f(100)| \leqslant e^{100}|f(0)|$.

48. Бесконечно дифференцируемая функция f на прямой такова, что $f(x) + f(2x) + f(3x) = 0$ при всех x. Можно ли утверждать, что f тождественно равна 0?

49. Функция f дважды непрерывно дифференцируема на прямой. Доказать, что предел $\lim_{h\to 0}[f(x+h)+f(x-h)-2f(x)]/h^2$ существует и равен $f''(x)$.

50. Может ли бесконечно дифференцируемая на прямой функция иметь бесконечно много корней на интервале $(0; 1)$ и не равняться 0 тождественно? Тот же вопрос для суммы степенного ряда (сходящегося при всех x).

51. Найти первообразные функций $1/(2 + 3x)$, $1/(x^2 + 3x + 2)$, $\ln x$.

52. Бесконечно дифференцируемая на прямой функция имеет период 1. Доказать, что предел среднего арифметического значений функции f в точках $0, 1/n, 2/n, \ldots, (n-1)/n$ равен $\int_0^1 f(t)\,dt$. Оценить скорость сходимости (ответ: $o(1/n^k)$ при любом k).

53. Построить последовательность бесконечно дифференцируемых функций f_k на $[-1;1]$ такую, что для любой непрерывной на этом отрезке функции g выполняется равенство $g(0) = \lim_{k\to\infty} \int_{-1}^{1} f_k(t)g(t)\,dt$.

54. Ряд $\sum a_i$ с положительными членами расходится, S_i — последовательность его частичных сумм. (а) Доказать, что ряд $\sum a_i/S_i$ расходится. (б) Доказать, что ряд $\sum a_i/S_i^2$ сходится.

55. Конечна ли сумма $1/n$ по всем натуральным n, десятичная запись которых не содержит восьмёрок?

56. Члены ряда $\sum a_i$ неотрицательны и убывают. Доказать, что он сходится или расходится одновременно с рядом $a_1 + 2a_2 + 4a_4 + 8a_8 + \ldots$

57. Конечны ли суммы (а) $1/(m^2+n^2)$, (б) $1/\mathrm{НОК}(m,n)$ по всем парам натуральных чисел m, n?

58. Доказать, что существуют такие A, B, C, что $1 + 1/2 + 1/3 + \ldots + 1/n = C + \ln n + A/n + B/n^2 + o(1/n^2)$. Найти A, B.

59. Конечна ли сумма $1/p$ по всем простым p?

60. Доказать, что $\int_x^{x+1} \sin t^2\,dt < 2/x$ при $x > 0$.

61. Если $f((x+y)/2) \leqslant (f(x)+f(y))/2$ для непрерывной функции f, то эта функция выпукла.

62. Доказать, что выпуклая функция дифференцируема всюду, кроме счётного множества точек.

6.3. Геометрия

1. Доказать, что взаимнооднозначное отображение координатной плоскости в себя линейно тогда и только тогда, когда оно переводит 0 в 0 и любую прямую — в прямую.

2. На какие отрезки делят диагональ n-мерного куба проекции его вершин?

3. В $\mathbb{R}^k$ имеются n векторов, все углы между которыми — тупые. Доказать, что $n \leqslant k+1$.

4. Сфера стереографически проектируется на плоскость и при этом может вращаться вокруг неподвижного центра. Возникают отображения плоскости в себя. Доказать, что это — дробно-линейные преобразования.

5. Если у ограниченной фигуры есть несколько осей симметрии, то все они пересекаются в одной точке.

6. Описать все конечные подгруппы группы подобий плоскости.

7. Найти явную формулу для чисел Фибоначчи.

8. Векторы $e_1, \ldots, e_n$ образуют базис евклидова пространства, $a_1, \ldots, a_n$ — произвольные числа. Доказать, что существует вектор x, для которого $(x, e_i) = a_i$.

9. Доказать, что в треугольнике со сторонами a, b, c и радиусом описанной окружности R расстояние между центром описанной окружности и ортоцентром равно $\sqrt{9R^2 - a^2 - b^2 - c^2}$.

10. В последовательности векторов скалярный квадрат каждого равен 2, а скалярное произведение соседних, а также первого и последнего, равно -1. Могут ли векторы быть линейно независимы?

11. При каких условиях композиция трёх поворотов плоскости с заданными центрами и углами равна тождественному преобразованию?

12. Построить окружность, проходящую через две данные точки и пересекающую данную окружность под данным углом.

13. На катетах треугольника построены квадраты. Доказать, что центры этих квадратов и середина гипотенузы являются вершинами прямоугольного равнобедренного треугольника.

14. Найти расстояние от точки n-мерного пространства до гиперплоскости в этом пространстве, если точка задана координатами, а гиперплоскость — уравнением.

15. Доказать, что сумма косинусов двугранных углов тетраэдра не больше 2.

16. (а) Найти все дробно-линейные преобразования, отображающие единичный круг $|z| < 1$ на себя. (б) То же для полуплоскости $\operatorname{Im} z > 0$.

17. Сопоставим каждому дробно-линейному преобразованию $z \mapsto \dfrac{az + b}{cz + d}$ матрицу $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$. Доказать, что композиции преобразований соответствует произведение матриц.

18. Доказать, что если некоторое подпространство в $\mathbb{R}^n$ представлено в виде объединения конечного числа подпространств, то оно совпадает с одним из них.

19. Доказать, что выпуклая оболочка любого конечного множества в $\mathbb{R}^n$ может быть представлена как пересечение конечного числа гиперплоскостей.

20. Доказать, что при преобразовании плоскости $\langle x, y\rangle \mapsto \langle x + y, y\rangle$ круг переходит в фигуру, имеющую оси симметрии, и найти их.

21. Найти ортогональную проекцию вектора $f : x \mapsto x^2$ в пространстве функций со скалярным произведением $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$ на плоскость, порождённую векторами x, $\sin x$.

22. Вокруг данного эллипсоида описываются всевозможные прямоугольные параллелепипеды. Доказать, что диагонали всех этих параллелепипедов равны.

23. Найти последовательность многочленов степени $0, 1, \ldots, n$, ортогональных относительно скалярного произведения $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$
(а) при $n = 4$; (б) при произвольном n.

24. При каких условиях заданный эллипс может быть получен как сечение заданного эллипсоида?

25. Доказать, что пятая по величине (считая от максимальной) ось n-мерного эллипсоида есть максимум малой оси всех пятимерных эллипсоидов, являющихся его сечениями.

26. Центры правильных треугольников, построенных на сторонах произвольного треугольника, образуют правильный треугольник.

27. Доказать, что площадь треугольного сечения тетраэдра не превосходит площади некоторой из его граней.

28. Три корня кубического многочлена с комплексными коэффициентами образуют треугольник. Доказать, что в этот треугольник можно вписать эллипс, фокусы которого находятся в корнях производной этого многочлена.

29. Три вектора в пространстве образуют друг с другом углы A, B, C. Доказать, что $\cos A + \cos B + \cos C \geqslant -3/2$.

30. Доказать, что расстояние d между центрами вписанной в треугольник окружности и описанной около него окружности связано с их радиусами r и R соотношением $d^2 = R^2 - 2Rr$.

Задачный раздел

В этом разделе вниманию читателей предлагается подборка задач разной степени сложности, в основном трудных. Некоторые из этих задач (не обязательно самые сложные!) требуют знания «неэлементарной» математики — анализа, линейной алгебры и т. п.

Составителям этой подборки кажется, что предлагаемые ниже задачи окажутся интересными как для сильных школьников, интересующихся математикой, так и для студентов-математиков.

Помимо *решения* задач, в высшей степени полезно упражняться в *изложении решений*. Мы советуем всем, решившим какую-либо из задач, постараться записать её решение в максимально простом и понятном виде и прислать в редакцию. В последующих номерах мы опубликуем самые изящные из полученных решений.

К сожалению, нам известны авторы далеко не всех предлагаемых ниже задач. Многие из них известны десятилетиями и стали частью «математического фольклора». Одна из целей, преследуемых составителями данного раздела, — записать этот «фольклор», многие части которого стремительно исчезают в наше время.

Мы обращаемся с просьбой ко всем читателям, имеющим свои собственные подборки таких задач, присылать их в редакцию. И, разумеется, мы с удовольствием будем публиковать свежие авторские задачи. Ждем ваших писем.

В скобках после условия задачи приводится фамилия автора (уточнения со стороны читателей приветствуются). Если автор задачи неизвестен, мы указываем того, кто предложил эту задачу.

1. Дана возрастающая функция $f(x)$ такая, что $f(0) > 0$, $f(1) < 1$. Докажите, что существует такое x, что $f(x) = x$ и, кроме того, x — точка непрерывности функции f.

2. Пусть $P(x)$ и $Q(x)$ — многочлены, причем $Q(0) = 0$. Докажите, что если $P(Q(x))$ — чётная функция, то $Q(x)$ — чётная или нечётная функция.

   (О. Ф. Крижановский)

3. Пусть $a_0 = a$, $a_{n+1} = a^{a_n}$, q — произвольное натуральное число, большее 1. Докажите, что последовательность остатков от деления a_n на q стабилизируется (т. е. все остатки, начиная с некоторого, равны).

4. Можно ли числа от 1 до 2^{1000} раскрасить в два цвета так, чтобы не существовало арифметических прогрессий длины 2000, составленных из чисел одного цвета?

5. Дано выпуклое тело в пространстве. Докажите, что можно отметить 4 точки на его поверхности так, чтобы касательная (т. е. опорная плоскость) в каждой отмеченной точке была параллельна плоскости, проходящей через остальные три. *(А. Я. Белов)*

6. Из произвольной точки P вне эллипса проведены два касательных к эллипсу луча l_1 и l_2. Кроме того, из P проведены лучи s_1 и s_2 через фокусы эллипса. Докажите, что угол между l_1 и s_1 равен углу между l_2 и s_2. *(В. В. Произволов)*

7. Конечно или бесконечно множество многочленов без кратных корней, со старшим коэффициентом 1, все коэффициенты которых целые, а все корни вещественны и принадлежат отрезку $[-1{,}99; +1{,}99]$? *(А. Я. Канель)*

8. Выяснить, равномерно ли сходится на отрезке $[0; 1]$ ряд

$$\sum \frac{x^n}{(1+x^n)^n}.$$

(А. Д. Соловьёв)

9. Даны матрицы $A_1, \ldots, A_k$ размера $n \times n$. Известно, что все произведения вида $A_{i_1} \cdot \ldots \cdot A_{i_h}$, где $h \leqslant n$ нильпотентны. Докажите, что любое произведение вида $A_{i_1} \cdot \ldots \cdot A_{i_{n^2}}$ равно нулю. *(И. П. Шестаков и И. В. Львов)*

10. Внутри выпуклого четырёхугольника $ABCD$ взята такая точка O, что $\angle AOP = \angle COQ$, где P и Q — точки пересечения продолжений сторон AB, CD и BC, AD соответственно. Докажите, что биссектрисы углов AOC и BCD перпендикулярны друг другу. *(С.Маркелов)*

Новые издания

Московский Центр непрерывного математического образования и Высший Математический Колледж Независимого Московского университета активно издают учебную и научную литературу по математике. Здесь приводится информация об изданиях МЦНМО и МК НМУ, вышедших во второй половине 1997 года и начале 1998 года, а также об изданиях, которые должны появиться в ближайшее время.

Издания МЦНМО

В. В. Прасолов, А. Б. Сосинский. ***Узлы, зацепления, косы и трехмерные многообразия.***

Эта книга, прежде всего, — введение в замечательные результаты Вогана Джонса и Виктора Васильева об инвариантах узлов и зацеплений и в новые модификации этих инвариантов, включая математическое обоснование инвариантов Джонса – Виттена. Особое внимание уделяется геометрическим аспектам теории. Обсуждаются такие темы, как косы, гомеоморфизмы поверхностей, перестройки трехмерных многообразий (исчисление Кирби), разветвленные накрытия. В двух последних главах строго математически строятся инварианты Джонса – Виттена на основе скейн-алгебр.

В отличие от недавних (зарубежных) монографий, в которых эти инварианты строятся на основе далеко продвинутых математических теорий (квантовые группы, теория представлений), в этой книге от читателя требуется минимальная математическая подготовка.

Многочисленные рисунки помогают яснее представить себе излагаемые геометрические конструкции. Изложение сопровождается задачами, которые позволяют использовать книгу в качестве учебника.

Для научных работников — математиков и физиков-теоретиков. Может быть использована аспирантами и студентами соответствующих специальностей.

В. В. Прасолов, В. М. Тихомиров. ***Геометрия.***

В книге дается систематическое изложение различных геометрий — евклидовой, аффинной, проективной, эллиптической, гиперболической, бесконечномерной. Проблемы различных геометрий рассматриваются с единой точки зрения, и всюду прослеживаются единые корни различных явлений. Все геометрические объекты исследуются с позиций двойственности. Подробно изложена теория коник и квадрик, в том числе и теория коник для неевклидовых геометрий. В книге изложено много ярких геометрических фактов, решено множество красивых геометрических задач.

Многочисленные рисунки помогают яснее представить себе излагаемые геометрические теоремы. В конце глав приводятся задачи и упражнения, которые позволяют использовать книгу в качестве учебника.

Книга призвана способствовать развитию геометрических исследований и совершенствованию математического образования. Для студентов, аспирантов, учителей математики, научных работников — математиков и физиков.

Готовится к изданию

Алгоритмы: построение и анализ.

Перевод книги *Introduction to Algorithms* (T. Cormen, C. Leicerson, R. Rivest). Классический учебник, содержащий наиболее важные методы построения эффективных алгоритмов. С момента выхода в США (1991) в издательстве M.I.T. Press выдержал более 10 изданий. Для программистов и интересующихся математиков.

Студенческие чтения МК НМУ

В этой серии предполагается публикация материалов лекций, прочитанных ведущими современными математиками для школьников и студентов.

В. И. Арнольд. ***Таинственные математические троицы и Принцип топологической экономии в алгебраической геометрии.***

Готовятся к изданию

Ю. И. Манин. ***Рациональные кривые, эллиптические кривые и уравнение Пенлеве.***

А. А. Кириллов. ***Метод орбит и конечные группы.***

Готовятся к изданию лекционные курсы МК НМУ

Ю. М. Бурман, Б. Л. Фейгин. ***Бесконечномерные алгебры Ли – I.***

Книга содержит описание представлений разнообразных бесконечномерных алгебр Ли (алгебры Гейзенберга, алгебры Вирасоро, алгебры полубесконечных матриц и др.) в пространстве полубесконечных форм. Описываются и изучаются вертексные операторы и соотношения между ними. Излагаются приложения этой теории к задачам алгебры и комбинаторики.

И. М. Парамонова, О. К. Шейнман. ***Задачи семинара «Группы и алгебры Ли».***

А. А. Белавин. ***Теоретическая физика.***

С. М. Натанзон. ***Геометрия двумерных топологических теорий поля.***

Опечатки, замеченные в №1

Страница,	Строка	Напечатано	Следует читать
8,	1 снизу	20-литровый	30-литровый
53,	7 сверху	$a_1, \ldots, a_{n+1}$	$c_1, \ldots, c_{n+1}$
53,	8 сверху	$= a_n(X - a_1) \ldots (X - a_n)$	$= a_n(X - c_1) \cdot \ldots \cdot (X - c_n)$
53,	8 сверху	$p(a_{n+1}) = 0 \neq q(a_{n+1})$	$p(c_{n+1}) = 0 \neq q(c_{n+1})$
53,	10 сверху	$a_1, \ldots, a_n$	$c_1, \ldots, c_n$
56,	3 сверху	$+ \sin \varphi)$	$+i \sin \varphi)$
56,	3 сверху	$z^n = R^n(\cos n\varphi + \sin n\varphi)$	$z^n = R^n(\cos n\varphi + i \sin n\varphi)$
60,	4 сверху	$e_f(x) > 0$	$e_f(x) > 1$
60,	14 снизу	$f : X \setminus D$	$f : X \setminus D \to \mathbb{C}$
65,	7 снизу	$\int_{\lvert z\rvert=R} \frac{dz}{zp(z)} = 2\pi i$	$\int_{\lvert z\rvert=R} \frac{dz}{zp(z)} = \frac{2\pi i}{p(0)}$
65,	6 снизу	любого	для любого
68,	15 снизу	Лагранжа	Лагранжем
122,	3 сверху	Вас	вас
138,	10 сверху	$\beta \leqslant 4$	$\beta < 4$
138,	13 сверху	$2\beta \leqslant a + c - 1$	$2b \leqslant a + c - 1$
154,	3 сверху	Если α — периодическое рациональное	Если многочлен $P(x) = x^2 - \alpha$ периодический и α — рациональное
161,	1 снизу	$k_1^2(x)$	$k_1(x)^2$
162,	1 сверху	$k_n^2(x)$	$k_n(x)^2$

Издательство Московского Центра
непрерывного математического образования

Технический редактор М. Н. Вялый

Лицензия ЛР №071150 от 11.04.95 г.
Подписано в печать 16.02.98 г. Формат 70×100/16
Печать офсетная. Печ. л. 14,0.
Тираж 1000. Заказ

МЦНМО
121002, Москва, Б. Власьевский, 11.