В. М. Тихомиров

РАССКАЗЫ О МАКСИМУМАХ И МИНИМУМАХ

Издание второе, исправленное

Москва
Издательство МЦНМО
2006

УДК 51(023)
ББК 22.10
Т46

Тихомиров В. М.

Т46 Рассказы о максимумах и минимумах. — 2-е изд., исправленное. — М.: МЦНМО, 2006. — 200 с.

ISBN 5-94057-250-2

Прослеживается история методов нахождения наименьших и наибольших величин. Подробно излагаются решения многих замечательных задач на максимум и минимум, принадлежащие великим математикам прошлых эпох — Евклиду, Архимеду, Герону, Тарталье, Ферма, Кеплеру, Бернулли, Ньютону и др. Говорится о зарождении многих идей, заложивших основания современного анализа. Объясняются связи экстремальных задач с проблемами естествознания, техники и экономики, рассказывается об основных принципах современной теории экстремальных задач, на основе теории экстремумов приводятся решения многих задач алгебры, геометрии, анализа.

Для школьников, учителей, студентов, преподавателей.

ББК 22.10

Тихомиров Владимир Михайлович

РАССКАЗЫ О МАКСИМУМАХ И МИНИМУМАХ

Редактор *Коробкова Т. Л.*
Корректор *Ботова А. С.*

Подписано в печать 07.07.06 г. Формат 60 × 90 1/16.
Бумага офсетная. Печать офсетная. Печ. л. 12,5.
Тираж 3000 экз. Заказ №

Издательство Московского центра непрерывного математического образования
119002, Москва, Большой Власьевский пер., 11. Тел. 241-74-83.

Отпечатано с готовых диапозитивов в ФГУП «Полиграфические ресурсы».

ISBN 5-94057-250-2

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ

В жизни постоянно приходится сталкиваться с необходимостью принять наилучшее возможное (иногда говорят — оптимальное) решение. Огромное число подобных проблем возникает в экономике и технике. При этом часто случается так, что полезно прибегнуть к математике.

В математике исследование задач на экстремум (т. е. на максимум и минимум началось очень давно — двадцать пять веков назад. Долгое время к задачам на отыскание экстремумов не было сколько-нибудь единых подходов. Но примерно триста лет назад — в эпоху формирования математического анализа — были созданы первые общие методы решения и исследования задач на экстремум.

Тогда же выяснилось, что некоторые специальные задачи оптимизации играют очень важную роль в естествознании, а именно обнаружилось, что многие законы природы допускают вывод из так называемых «вариационных принципов», согласно которым истинное движение механической системы, света, электричества, жидкости, газа и т. п. можно выделить из произвольной совокупности допустимых движений тем, что они минимизируют или максимизируют некоторые величины. В конце XVII столетия было поставлено несколько конкретных экстремальных задач естественнонаучного содержания (брахистохрона, задача Ньютона и др.). Потребность решать как их, так и многие другие проблемы, возникающие в геометрии, физике, механике, привела к созданию новой главы математического анализа, получившей название вариационного исчисления.

Интенсивное развитие вариационного исчисления продолжалось около двух столетий. В нем принимали участие многие замечательные ученые XVIII и XIX веков, и к началу XX столетия стало казаться, что они почти исчерпали эту тематику.

Но это оказалось не так. Потребности практической жизни, особенно в области экономики и техники, в последнее время выдвинули такие новые задачи, которые старыми методами решить не удавалось. Надо было идти дальше. Пришлось несколько развить математический анализ и создать новый его раздел — «выпуклый анализ», где изучались выпуклые функции и выпуклые экстремальные задачи.

С другой стороны, потребности техники, в частности космической, выдвинули серию задач, которые также не поддавались средствам вариационного исчисления. Необходимость решать их привела к созданию новой теории, получившей название теории оптимального управления. Основной метод в теории оптимального управления был разработан в 1950—60-е годы нашими соотечественниками — Л. С. Понтрягиным и его учениками. Это привело к тому, что теория экстремальных задач получила новый мощный толчок к дальнейшим исследованиям.

Цель данной книги — познакомить читателя со всем этим кругом идей. Но эта цель — не единственная. Задачи на максимум и минимум на протяжении всей истории математики играли важную роль в развитии этой науки. За все это время накопилось большое число красивых, важных, ярких и интересных задач в геометрии, алгебре, физике и т. п. В решении этих конкретных задач принимали участие крупнейшие ученые прошлых эпох — Евклид, Архимед, Аполлоний, Герон, Тарталья, Торричелли, Иоганн и Якоб Бернулли, Ньютон и многие другие. Решение конкретных задач стимулировало развитие теории, и в итоге были выработаны приемы, позволяющие единым методом решать задачи самой разнообразной природы.

Хочется, чтобы читатель понял, как и зачем рождается математическая теория. В первой части книги он познакомится со многими конкретными задачами, при обсуждении их решений соприкоснется с творчеством ряда крупнейших математиков прошлого. И это имеет не только исторический интерес. Идеи и методы, созданные замечательными математиками при решении конкретных проблем, обычно не умирают. Они потом обязательно где-нибудь возрождаются, и потому проникновение в замыслы великих людей всегда обогащает.

Но когда возникает необходимость в решении большого числа разнообразных проблем, создаются предпосылки для создания общей теории. Во второй части рассказывается об одном методе решения задач на максимум и минимум, восходящем к Лагранжу. Основной замысел этого метода сохраняется на протяжении более двух столетий. Меняется его наполнение, но центральная мысль при этом остается неизменной. Понять причины такой универсальности идеи Лагранжа непросто. Однако научиться пользоваться лагранжевским принципом для решения задач совсем не трудно, и в конце второй части книги все задачи, которые обсуждались вначале, задачи, решения которых были столь непохожи друг на друга, исследуются нами с помощью единого общего приема, стандартно, с применением одной и той же схемы.

Автор старался показать, как из анализа разрозненных фактов рождается общая идея, как она трансформируется, обогащается новым содержанием и вместе с тем остается сама собой.

Все содержание книги, кроме заключительной части четырнадцатого рассказа, адресовано прежде всего школьнику. Но мне хотелось бы видеть среди читателей книги и студентов, интересующихся математикой и, конечно, учителей. Последний рассказ адресован прежде всего им. Там затрагивается вопрос о том, как и чему следует учить. Содержание книги, мне думается, дает благодарный материал для того, чтобы обсудить эту тему, которая еще очень долго будет всех волновать. Я надеюсь также, что эту книгу прочтут и мои коллеги, занимающиеся математикой и преподающие ее студентам.

Хочу выразить благодарность тем, кто читал рукопись и высказал мне свои замечания о ней. Прежде всего это относится к Андрею Николаевичу Колмогорову, Николаю Борисовичу Васильеву, Ивану Пенкову и Георгию Георгиевичу Магарил-Ильяеву.

P. S. Это предисловие было написано в начале 1986 года. Оставляю его без изменений. Скажу лишь с глубокой горечью, что прошедшие два десятка лет унесли из жизни моего дорогого учителя А. Н. Колмогорова и светлого человека Н. Б. Васильева. Благодарю А. Шмуклер, указавшую на многие опечатки в первом издании.

1 июля 2006

В. Тихомиров

ЧАСТЬ ПЕРВАЯ

СТАРИННЫЕ ЗАДАЧИ НА МАКСИМУМ И МИНИМУМ

Математика владеет не только истиной, но и высшей красотой,... доступной только величайшему искусству.

Б. Рассел

Следовать за мыслями великого человека есть наука самая занимательная.

А. С. Пушкин

Рассказ первый

ЗАЧЕМ РЕШАЮТ ЗАДАЧИ НА МАКСИМУМ И МИНИМУМ?

В мире не происходит ничего, в чем бы не был виден смысл какого-нибудь максимума или минимума.

Л. Эйлер

Большая часть вопросов практики приводится к задачам наибольших и наименьших величин,... и только решением этих задач мы можем удовлетворить требованиям практики, которая везде ищет самого лучшего, самого выгодного.

П. Л. Чебышёв

...Хочется дойти до самой сути.

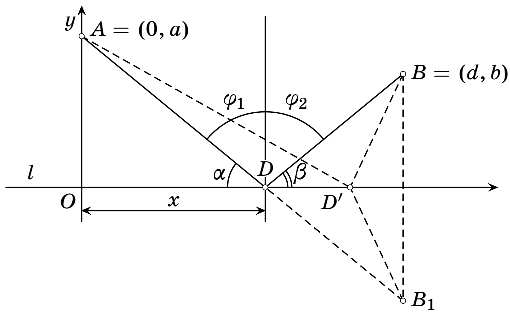
Б. Л. Пастернак

О максимумах и минимумах мы узнаем в школе. Вот одна старинная задача, которую вы могли решать на уроках геометрии.

Даны две точки A и B по одну сторону от прямой l. Требуется найти на l такую точку D, чтобы сумма расстояний от A до D и от B до D была наименьшей (рис. 1).

Здесь надо найти наименьшее значение, т. е. *минимум*. Во многих задачах требуется найти *максимум*, т. е. наибольшее значение чего-нибудь. Оба понятия — максимум и минимум —

объединяются единым термином — «*экстремум*», что по-латыни означает «крайнее». Задачи на отыскание максимума и минимума называются *экстремальными задачами*. (Почти тот же самый смысл вкладывается в название «*задачи оптимизации*».) Методы решения и исследования различного рода экстремальных задач составляют специальные разделы математического анализа. Они объединяются в общую главу, называемую *теорией экстремальных задач*.



Здесь наша цель — обсудить два вопроса: зачем решают задачи на максимум и минимум и из каких компонентов складывается теория экстремальных задач.

Выше была поставлена геометрическая задача. Ее можно встретить почти в каждом учебнике геометрии. Когда же она появилась впервые? И зачем?

Считают, что автором этой задачи является известный математик античности Герон Александрийский. (Далее мы называем ее *задачей Герона*. О Героне мы все знаем благодаря формуле площади треугольника, носящей его имя.) Книга, где была помещена эта задача, называлась «О зеркалах». О времени ее написания идут споры, но большинство исследователей сходятся на том, что она написана в I веке до н. э. Сам труд Герона не сохранился, и о нем известно из комментариев к нему, написанных позже. Думаю, что читатель уже слышал про задачу Герона и решал ее. Я привел ее потому, что нам она будет очень полезна для разного рода иллюстраций.

Вспомним, как решается задача Герона.

Пусть B_1 — точка, симметричная B относительно прямой l. Соединим A с B_1. Тогда *точка D пересечения AB_1 с прямой l будет искомой* (см. рис. 1).

Действительно, для любой точки D', отличной от D, имеет место неравенство

$$|AD'| + |D'B| = |AD'| + |D'B_1| > |AB_1| = |AD| + |DB_1|. \qquad (1)$$

Здесь и в дальнейшем мы будем использовать следующие обозначения: $[AB]$ — отрезок, соединяющий точки A и B, $|AB|$ — длина отрезка $[AB]$, $AB \parallel CD$ — прямая AB параллельна прямой CD.

В неравенстве (1) мы использовали свойства симметрии, из которых следуют равенства $|DB| = |DB_1|$, $|D'B| = |D'B_1|$ и неравенство треугольника $|AD'| + |D'B_1| > |AB_1|$. Задача решена.

Отметим: искомая точка D обладает тем свойством, что угол α равен углу β (см. рис. 1), а также угол φ_1 равен углу φ_2 или, как говорят, *угол падения равен углу отражения*.

Попробуйте теперь, используя идею, заложенную в только что проведенном рассуждении, решить следующие задачи.

Задача 1. *Дан угол и точка C внутри него. Найти точки A и B на сторонах угла так, чтобы периметр треугольника ABC был наименьшим.*

Задача 2. *Дан угол и две точки C и D внутри него. Найти точки A и B на сторонах угла так, чтобы сумма длин* $|CA| + |AB| + |BD|$ *была наименьшей.*

Но вернемся к задаче Герона. Герон исследует в своей книге законы отражения света и прилагает итоги своих размышлений к вопросам, связанным со свойствами зеркал. В частности, он доказывает, что параболическое зеркало фокусирует пучок лучей, параллельных оси параболы.

В ту пору законы природы старались постичь умозрительно, с помощью логических рассуждений, не прибегая к эксперименту. У нас еще будет повод поговорить о зарождении современной науки, опирающейся на опыт. Первым экспериментатором в истории науки был Г. Галилей, который жил в XVII веке, Герон же при объяснении законов отражения искал для них логические основания. Он высказал предположение, что природа действует кратчайшим путем. Вот как пишет об этом один из его комментаторов Дамианос (VI век н. э.): «Герон ... показал, что прямые, наклоненные под равными углами, являются самыми меньшими из всех промежуточных, образующих наклоны с одной и той же стороны от одной и той же прямой. Доказывая это, он говорит, что если природа не хочет попусту обводить луч зрения, то она изломит их под равными углами».

Исследователи истории науки полагают, что здесь впервые прозвучала мысль о том, что природа управляется *экстремальными принципами*. Идею Герона развил Ферма. (На этом мы подробнее остановимся дальше — в третьем рассказе.) Ферма вывел известный к тому времени закон преломления света, исходя из допущения, что траектория распространения света от одной точки до другой в неоднородной среде характеризуется тем, что вдоль нее тратится наименьшее время. Начиная с этого момента, идея экстремальности проявлений природы становится путеводной звездой всего естествознания. В подтверждение и были приведены в качестве эпиграфа слова Эйлера. Отложим пока обсуждение удивительности этого феномена, но факт остается фактом: траектории света и радиоволн, движения маятников и планет, течения жидкостей и газов и многие другие движения выделяются из многообразия всех возможных движений тем, что они являются решениями некоторых задач на максимум или минимум. Это обстоятельство оказывается плодотворным средством математического описания природы.

Вот в чем состоит первая причина, побуждающая решать задачи на максимум и минимумmp иmp развиватьmp теориюmp экстремальных задач. Она привела в XVIII веке к созданию специального раздела этой теории, который был назван *вариационным исчислением*.

Вторая причина кроется в нас самих. Людям свойственно стремление к лучшему, и потому им всегда хочется выбрать оптимальную из имеющихся возможностей. Случается, что математика может здесь помочь.

Обсудим это, и пусть снова задача Герона послужит нам примером. Иногда в учебниках ей придают вид проблемы, возникающей на практике. Тогда прямая l становится, скажем, прямолинейным участком железной дороги, точки A и B — городами, точка D — железнодорожной платформой. И ставится вопрос: *где следует поставить платформу, чтобы соединяющие ее с городами прямолинейные шоссейные участки имели наименьшую суммарную длину?*

Вот еще несколько геометрических задач, которые могут иметь прикладное значение. Попробуйте продумать их самостоятельно.

Задача 3. *Пусть имеются три города — A, B и C. Требуется указать такое место D, чтобы суммарная длина прямолинейных участков шоссе, соединяющих D с A, B и C, была минимальной.*

Задача 4. *То же самое, что и в задаче 3, но для четырех* *городов.*

Задача 5. *Изменится ли решение предыдущей задачи, если поставить вопрос о наименьшей длине шоссейной дороги, соединяющей 4 города, и при этом не уточнено, что все пути должны соединяться в одной точке*?

Конечно, такого рода задачи представляют собой лишь модели реальных жизненных ситуаций. На практике все обстоит гораздо сложнее: и участки железной дороги не бывают прямолинейными, и шоссе не строят строго по прямым, и «сумма расстояний» в чистом виде редко бывает «критерием оптимальности». Но несомненно, что и при строительстве железных, шоссейных или иных дорог, равно как и при строительстве газо- и нефтепроводов, и при многом другом, обычно возникает вопрос о том, как это осуществить наиболее целесообразно, скажем, наиболее дешево.

Такие проблемы постоянно возникают в хозяйственной деятельности. Все время приходится изыскивать или самый дешевый, или самый быстрый, или самый короткий, или самый экономичный способ достижения цели.

Приведем пример проблемы оптимизации экономического содержания. Пусть имеются базы с некоторым продуктом, магазины и автопарк. Как следует диспетчеру автопарка организовать доставку необходимого продукта в магазины наиболее экономично? (Такого типа задачи называются «транспортными». Впоследствии мы уточним их постановку.) При решении сходных вопросов приходится обращаться к математике. Методы, разработанные для решения задач на максимум и минимум к середине XX века, оказались недостаточными для решения подобных проблем.

Выяснилось, в частности, что во многих задачах экономического содержания большую роль играет понятие *выпуклости* — там часто встречаются выпуклые и даже линейные функции и множества. Пришлось существенно развить теорию выпуклых множеств и функций. Эта теория получила даже специальное название — *выпуклый анализ*. Были созданы и новые направления в теории экстремальных задач — *линейное* и *выпуклое программирование*. Начала этих направлений были заложены в 1930-е годы советским математиком Л. В. Канторовичем.

Большое число задач оптимизации возникает в технике. Это задачи управления технологическими процессами, приборами, системами. Вот пример. Пусть имеется тележка, движущаяся прямолинейно без трения по горизонтальным рельсам. Тележка управляется внешней силой, которую можно изменять в заданных пределах. Требуется остановить тележку в определенном положении в кратчайшее время. Эта задача называется *простейшей задачей о быстродействии в автоматическом регулировании*. Очень много задач возникло в химической промышленности, в космонавтике и т. п. При этом выяснилось также, что методы вариационного исчисления недостаточны для решения этих задач. Пришлось создавать новую главу, дополняющую вариационное исчисление. Она получила название *оптимального управления*.

Вот и вторая причина, заставляющая нас решать задачи оптимизации и развивать теорию экстремальных задач, — желание «удовлетворить требованиям практики», о котором говорил Чебышёв.

Но этими двумя причинами нельзя объяснить всего.

В следующем рассказе речь пойдет о древнейшей задаче на максимум и минимум — о классической изопериметрической задаче. Около двадцати пяти веков назад в Древней Греции было открыто замечательное свойство круга — среди замкнутых кривых заданной длины охватывать наибольшую площадь. Вы, наверное, решали в школе задачи, описывающие аналогичные свойства многоугольников. Вспомним две из них.

Задача 6. *Найти треугольник заданного периметра, имеющий наибольшую площадь.*

Задача 7. *Доказать, что квадрат имеет наибольшую площадь среди всех прямоугольников с заданным периметром.*

Вопрос, равносильный последней задаче, рассматривался еще в «Началах» Евклида; решением этой же задачи Ферма проиллюстрировал свой метод нахождения максимумов и минимумов, известный нам как теорема Ферма.

Зачем же ставились и для чего решались такие задачи? Что привлекает в них? Почему в большинстве книг по геометрии авторы так любят обсуждать задачи на максимум и минимум?

Это не так легко объяснить, но факт остается фактом, что на протяжении всей истории математики задачи на экстремум вызывали интерес и желание решать их. Может быть, все дело в том, что человеку свойственно стремление к совершенству, в том, что имеется какой-то таинственный стимул постижения «самой сути»? А может быть, в экстремальных задачах всегда или, по крайней мере, часто присутствует что-то изящное, привлекательное, нечто от той красоты, о которой говорит Рассел? И именно это побуждает нас решать задачи на максимум и минимум?

Сказанного достаточно, чтобы понять важность и увлекательность избранного нами предмета.

Но, может быть, нелишне сказать о временны́х границах наших рассказов. Первые задачи на максимум и минимум были поставлены в очень далекие времена: классическая изопериметрическая задача обсуждалась в V веке до н. э. О ней мы поговорим в следующем рассказе. А в предпоследнем, четырнадцатом, рассказе мы обсудим проблемы, которые возникают в наши дни.

Долгое время каждая задача на экстремум решалась индивидуально. В XVII веке явственно стала ощущаться необходимость создания каких-то общих методов. Такие методы были разработаны Ферма, Ньютоном, Лейбницем и другими — сначала для одной, потом для нескольких, а затем и бесконечного числа переменных. В итоге сформировались основные разделы теории экстремальных задач: *математическое программирование*, т. е. теория конечномерных задач оптимизации, *выпуклое* (в том числе линейное) *программирование* (где изучают выпуклые задачи оптимизации), *вариационное исчисление* и *теория оптимального управления*.

Эта книга состоит из двух частей. В первой собраны старинные задачи, поставленные и решенные, как правило, до того, как были созданы первые общие методы. Во второй части рассказывается о некоторых методах теории экстремальных задач.

В первой части обсуждаются задачи, связанные с именами крупнейших математиков разных времен — Евклида, Архимеда, Ферма, Кеплера, Гюйгенса, И. Бернулли, Ньютона, Лейбница. Я не отказал себе в удовольствии «следовать за мыслями» этих великих людей.

Во второй же части... Но об этом пока еще рано говорить.

Рассказ второй

ДРЕВНЕЙШАЯ ЗАДАЧА — ЗАДАЧА ДИДОНЫ

Столько купили земли и дали ей имя Бирса,
Сколько смогли окружить бычьей шкурой.

Вергилий. Энеида

Прекраснейшим телом является шар, а прекраснейшей плоской фигурой — круг.

Пифагор

Эпиграфом к нашему рассказу поставлены «из Энеиды два стиха» одного из величайших поэтов Древнего Рима — Публия Вергилия Марона. Как и всякое бессмертное творение, «Энеида» повествует о страстях человеческих, о добре и зле, о роке и страдании, о коварстве и любви, о жизни и смерти. Приведенные строки относятся к событию, произошедшему, если верить преданию, в IX веке до н. э. Вспомним легенду, воспроизведенную в «Энеиде».

Финикийская царевна Дидона, спасаясь от преследований своего брата, отправилась на запад вдоль берегов Средиземного моря искать себе прибежище. Ей приглянулось одно место на побережье нынешнего Тунисского залива. Дидона повела переговоры с местным предводителем Ярбом о продаже земли. Запросила она совсем немного — столько, сколько можно «окружить бычьей шкурой». Дидоне удалось уговорить Ярба. Сделка состоялась, и тогда Дидона изрезала шкуру быка на мелкие тесемки, связала их воедино и окружила большую территорию, на которой основала крепость, а вблизи от нее — город Карфаген. Там ожидали ее впоследствии неразделенная любовь и мученическая смерть.

Этот эпизод дает повод задуматься над вопросом: сколько же земли можно окружить бычьей шкурой?

Почему мы начинаем именно с этой задачи? Ведь решение ее довольно сложно. Казалось бы, следовало начинать с более простых вещей. Но я все-таки выбираю другой путь. Здесь, в этой части, я буду вести вас не от простого к сложному, а от далекого прошлого к нашим дням и потому начать хочу «с самого начала». Разве не удивительно, что в те «баснословные года» ставились и решались столь трудные и глубокие проблемы? Как мало знали наши предшественники в сравнении с тем, что знаем мы с вами! Но они шли к цели и достигали ее!

Итак, сколько же земли можно окружить бычьей шкурой? Для того чтобы ответить на этот вопрос, нужно правильно математически поставить задачу. Современный математик скажет так:

Среди замкнутых плоских кривых, имеющих заданную длину, найти кривую, охватывающую максимальную площадь.

Эту задачу и называют *задачей Дидоны* или *классической изопериметрической задачей*. (Изопериметрические фигуры — это фигуры, имеющие одинаковый периметр.)

Мы пока обошлись только словами, и человека с достаточным уровнем математической культуры такая формулировка вполне удовлетворит, ибо он знает, что такое кривая, длина и площадь. На придание же точного смысла этим словам ушло свыше двух тысяч лет. Чтобы как следует объяснить эти термины, следовало бы написать отдельную книгу. Не будем чрезмерно углубляться в это и подойдем к нашей задаче «наивно», как подходили к ней древние (и как должна была на практике подойти к ней сама царевна Дидона).

Попробуем только обойтись без шкуры быка. Отмотаем от катушки кусочек нити. Отрежем его и свяжем концами. Положим эту связанную нить на лист бумаги. Получилась *плоская замкнутая кривая*. Если теперь вырезать кусок бумаги по контуру нити, получится образ *площади, охватываемой этой кривой*. Эту площадь можно измерить. Измерение можно произвести достаточно точно, если наш лист был листом миллиметровки. Теперь уже можно понять и вопрос задачи: требуется выяснить, *как следует положить нашу нить, чтобы она охватывала наибольшую площадь*. Мы докажем чуть позже, что кривая, решающая классическую изопериметрическую задачу, — это окружность. Вергилий при описании действия Дидоны употребил глагол «circumdare» (окружать), содержащий корень «circus» (круг), что позволяет думать, что классическую изопериметрическую задачу сама Дидона решила правильно.

Из истории классической изопериметрической задачи. Многие историки полагают, что это — первая экстремальная задача, обсуждавшаяся в научной литературе. Вместе с изопериметрическим свойством круга (т. е. свойством окружности охватывать наибольшую площадь среди изопериметрических фигур) античные геометры отмечали *изопифанное свойство шара* (т. е. свойство сферы охватывать наибольший объем среди изопифанных фигур — фигур, имеющих равную площадь поверхности). С этим свойством — наибольшей вместимости — связаны представления о круге и шаре как воплощении геометрического совершенства (вспомним слова Пифагора, взятые эпиграфом к этому рассказу). Вот еще одно подтверждение той же мысли. Великая книга Н. Коперника начинается со слов: «Прежде всего мы должны заметить, что мир является шарообразным или потому, что эта форма совершеннейшая из всех и не нуждается ни в каких скрепах, и вся представляет цельность, или потому, что эта форма среди всех других обладает наибольшей вместимостью, что более всего приличествует тому, что должно охватить и сохранить все».

Сейчас невозможно сказать, когда впервые была высказана мысль о наибольшей вместимости круга и шара. Во всяком случае, Аристотель (IV век до н. э.) — один из величайших мыслителей в истории человечества — пользуется этими фактами, как известными. А кто же (не считая Дидоны) на самом деле решил классическую изопериметрическую задачу? Литература, посвященная изопериметрическому свойству круга и изопифанному свойству шара, огромна. Из необозримого числа работ назовем одну книгу — монографию немецкого геометра В. Бляшке [4]. Там есть и указания исторического характера. Среди тех, кто дал решение изопериметрической и изопифанной задач, древние авторы называли и Архимеда. Считается, что первые *строгие* доказательства максимального свойства круга и шара дал Г. А. Шварц. Если вам захочется проследить «историю „изопериметрической задачи“, которая началась в седой древности легендой о карфагенской царице Дидоне, до господина тайного советника Германа Амандуса Шварца из Берлина», вы можете обратиться к статье Бляшке [5] (слова, заключенные выше в кавычки, взяты из [4]).

Но на самом деле Шварцу, а до него Вейерштрассу и после него — самому Бляшке, как и многим другим математикам XIX и XX столетий, принадлежит (в отношении классической изопериметрической задачи) лишь оформление идей своих далеких предшественников, оформление, способное удовлетворить современным требованиям строгости. Основные же пути решения изопериметрической задачи были абсолютно правильно намечены еще в античные времена. Сейчас мы расскажем об одном из таких путей, принадлежащем Зенодору — математику, жившему, как считают, где-то между III веком до н. э. и I веком н. э.

Зенодор совершенно строго, на современном уровне этого понятия, доказывает следующее утверждение.

Если существует плоский n-угольник, имеющий наибольшую площадь среди всех n-угольников с заданным периметром, то он должен быть равносторонним и равноугольным.

Плоский n-угольник, имеющий наибольшую площадь среди всех изопериметрических с ним n-угольников, будем называть (для краткости) *максимальным n-угольником*. Используя этот термин, теорему Зенодора можно сформулировать короче.

Максимальный n-угольник (если он существует) *должен быть правильным.*

Теорема Зенодора является следствием двух лемм.

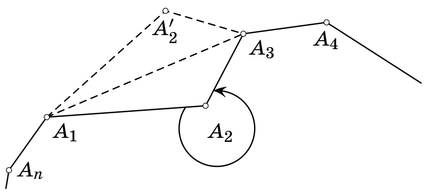
Лемма 1. *Максимальный n-угольник должен быть равносторонним.*

Лемма 2. *Максимальный n-угольник должен быть равноугольным.*

Излагая работы наших далеких предшественников, я не буду, как правило, воспроизводить их буквально, не стану сохранять обозначения и стиль авторов или стремиться приводить именно авторские доказательства. Мне хотелось бы воспроизвести лишь основное направление мысли и общий дух рассуждений, изменяя и модернизируя формулировки и доказательства. Так и здесь я приведу обработки доказательств лемм 1 и 2. При этом двукратно будет использовано решение задачи Герона.

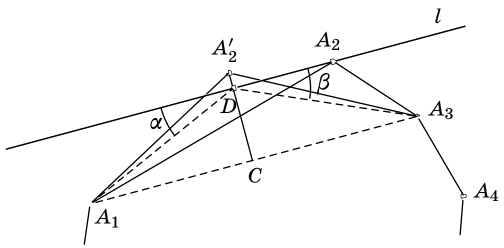
Прежде чем приступить к доказательствам, нужно сделать одно замечание.

Следует упомянуть об обстоятельстве, не оговоренном Зенодором. Покажем, что *невыпуклый многоугольник не может быть максимальным*. Действительно, если, скажем, угол $A_1A_2A_3$ больше 180° (рис. 2), то, отразив вершину A_2 относительно прямой A_1A_3 и рассмотрев многоугольник $A_1A_2'A_3\ldots A_n$, где A_2' —



образ A_2, мы получим изопериметрический многоугольник большей площади, чем площадь $A_1A_2\ldots A_n$. Теперь уже можно привести доказательства.

Доказательство леммы 1. Пусть $A_1A_2\ldots A_n$ — максимальный n-угольник. Тогда он, как было отмечено, является выпуклым. Допустим, что не все его стороны имеют одинаковые длины, и придём к противоречию. Пусть две какие-либо смежные стороны, скажем, A_1A_2 и A_2A_3, не равны между собой по длине. Через вершину A_2 проведём прямую l, параллельную A_1A_3 (рис. 3). Рассмотрим задачу Герона для прямой l и точек

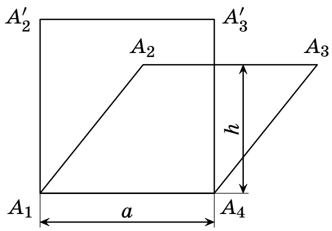


A_1 и A_3 об определении точки D на l, для которой сумма расстояний $|A_1D|+|A_3D|$ была бы минимальной. В предыдущем пункте было доказано, что в искомой точке D угол α равен углу β. Но угол α равен углу DA_1A_3, а угол β равен углу DA_3A_1 по свойству накрест лежащих углов при параллельных. Таким образом, треугольник A_1DA_3 — равнобедренный и, значит, точка D отлична от точки A_2. Вместе с тем:

а) площадь $\triangle A_1DA_3$ равна площади $\triangle A_1A_2A_3$, ибо у них одинаковые высоты и основания;

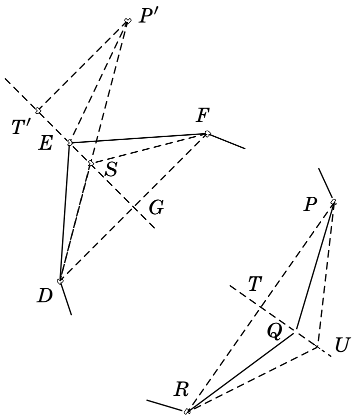
б) сумма боковых сторон треугольника A_1DA_3 меньше суммы сторон A_1A_2 и A_2A_3, ибо D ($\neq A_2$) есть решение задачи Герона. Построим теперь равнобедренный треугольник $A_1A_2'A_3$, у которого $|A_1A_2'|+|A_2'A_3|=|A_1A_2|+|A_2A_3|$. Его площадь, разумеется, больше площади $\triangle A_1A_2A_3$, так как высота $|A_2'C|$ больше высоты $|DC|$ (в силу того, что наклонная $|A_1A_2'|$ длиннее наклонной $|A_1D|$), значит, площадь многоугольника $A_1A_2'\ldots A_n$ больше площади изопериметрического с ним многоугольника $A_1A_2\ldots A_n$, что противоречит максимальности последнего многоугольника. Лемма 1 доказана.

С л е д с т в и е. Из леммы 1 вытекает, что максимальным треугольником является равносторонний и что максимальный четырехугольник должен быть ромбом. Из последнего заключения и из рис. 4 немедленно можно вывести, что максимальный четырехугольник на самом деле — квадрат.



Д о к а з а т е л ь с т в о л е м м ы 2. Пусть снова $A_1A_2\ldots A_n$ — максимальный n-угольник. Мы уже знаем (лемма 1), что все стороны его равны, и помним, что он выпуклый. Допустим теперь, что не все его углы одинаковы, и придем к противоречию. Если не все углы равны, то существуют два неравных смежных угла $\alpha \neq \beta$. Докажем, что тогда существуют и два неравных несмежных угла. Рассмотрим последовательно расположенные углы многоугольника α, β, γ, δ, ε, ... (их не меньше пяти). Если $\gamma \neq \alpha$ или $\delta \neq \beta$, то мы достигли цели, так как углы α и γ (или β и δ) — несмежные. Остается рассмотреть случай $\alpha = \gamma$, $\beta = \delta$, $\alpha \neq \beta$, когда последовательность имеет вид: α, β, α, β, ε, ... Но здесь не равны друг другу два несмежных угла — первый и четвертый.

Таким образом, при сделанном предположении можно считать, что существуют два непересекающиеся внутренними частями треугольника DEF и PQR (рис. 5), каждый из который образован подряд идущими вершинами нашего n-угольника, причем угол E меньше угла Q. В силу того, что $|DE| = |EF| = |PQ| = |QR|$, и из-за неравенства между углами, получаем, что $|DF| < |PR|$. Опустим из E и Q перпендикуляры EG на DF и QT на PR. К продолжению отрезка EG приложим треугольник $ET'P'$, равный (конгруэнтный) треугольнику QTP (точка T переходит в T', P — в P', Q — в E). И снова рассмотрим задачу Герона для прямой $T'G$ и точек P' и F. Пусть S — решение задачи Герона, т. е. точка на $T'G$, для которой сумма расстояний от P' до S и от S до F минимальна. В силу того, что угол $P'ET'$ (равный



половине угла Q) больше угла FEG (равного половине угла E), точка S не совпадает с точкой E (ибо углы $P'ST'$ и FSG равны) и, более того, S лежит на отрезке EG. Отложим теперь на прямой QT отрезок TU, равный по длине отрезку $T'S$, и рассмотрим треугольники DSF и PUR. Сумма боковых сторон этих треугольников меньше суммы боковых сторон исходных треугольников DEF и PQR. Действительно,

$$|DS| + |SF| + |PU| + |UR| = 2(|SF| + |SP'|) < 2(|FE| + |EP'|) = \\ = |DE| + |EF| + |PQ| + |QR|.$$

Мы пользовались тем, что наши треугольники равнобедренные, и тем, что S — решение задачи Герона. С другой стороны, площадь $\triangle P'ES$ больше площади $\triangle ESF$, ибо у первого высота равна $|P'T'| = \frac{1}{2}|PR|$, у второго высота равна $|FG| = \frac{1}{2}|DF|$, а по доказанному уже $|DF| < |PR|$. Отсюда следует, что сумма площадей треугольников DSF и PUR больше суммы площадей первоначальных треугольников DEF и PQR. Действительно,

$$S_{\triangle DSF} + S_{\triangle PUR} = S_{\triangle DEF} - 2S_{\triangle ESF} + S_{\triangle PQR} + S_{\triangle ESP'} > \\ > S_{\triangle DEF} + S_{\triangle PQR}. \quad (1)$$

Значит, n-угольник $DSF\ldots PUR\ldots$ имеет меньший периметр и большую площадь, чем наш изначальный n-угольник $DEF\ldots$ $\ldots PQR\ldots$ Теперь можно с любым из треугольников (DSF или POR) поступить точно так же, как мы поступили с $\triangle A_1DA_3$ при доказательстве леммы 1, т. е. надстроить его, уравняв периметры многоугольников и сделав площадь нового многоугольника еще большей, чем площадь многоугольника $DEF\ldots PQP\ldots$, следовательно, этот многоугольник — не максимальный. Лемма 2 полностью доказана, а с нею доказана и теорема Зенодора. Осталось вывести из нее решение классической изопериметрической задачи.

Лемма о существовании максимального n-угольника. Мы доказали, что если максимальный n-угольник существует, то он правильный. Но существует ли максимальный n-угольник? А вдруг нет? Тогда все пойдет прахом. А ведь не всякая функция имеет максимум. Например, функция $f(x) = -(1 + x^2)^{-1}$ не достигает своего наибольшего значения (подробнее этот пример разобран в одиннадцатом рассказе).

Вопросы существования решений не были предметом рассмотрения древних авторов. Значение проблем существования и методы доказательства теорем существования были поняты примерно сто лет назад. В дальнейшем нам придется не раз касаться этих вопросов. Здесь же мы приведем без доказательства следующее утверждение (которое Зенодор, по-видимому считал само собой разумеющимся).

Лемма 3. *Максимальный n-угольник существует.*

Отсюда из лемм 1 и 2 следует

Теорема 1. *Максимальный n-угольник является правильным n-угольником.*

Теперь осталось уже совсем немного.

Завершение доказательства. Пусть P — периметр правильного n-угольника, а S — его площадь. Мы знаем из геометрии, что $P = 2nR\sin(\pi/n)$, где R — радиус описанного круга, $S = rP/2$, где r — радиус вписанного круга. При этом $r = R\cos(\pi/n)$. Сопоставляя все это, приходим к формуле, связывающей S и P:

$$P^2 - 4n\,\mathrm{tg}\,\frac{\pi}{n}\cdot S = 0.$$

Теорема 1 означает, что если P — периметр некоторого произвольного n-угольника, а S — его площадь, то имеет место неравенство

$$P^2 - 4n\,\mathrm{tg}\,\frac{\pi}{n}\cdot S \geqslant 0. \qquad (2)$$

Из неравенства $\mathrm{tg}\,\alpha \geqslant \alpha$ (верного для $0 \leqslant \alpha < \pi/2$) и из (2) получим неравенство

$$P^2 - 4\pi S \geqslant 0, \qquad (3)$$

справедливое для любого n-угольника и любого n. Отметим, что для любого круга имеет место очевидное равенство

$$P^2 - 4\pi S = 0, \qquad (4)$$

где P — длина окружности, а S — площадь круга.

Теперь сформулируем лемму, связывающую все понятия, участвующие в формулировке классической изопериметрической задачи, с понятием n-угольника. Она означает, что *длину кривой и площадь, охватываемую ею*, можно с любой степенью точности приблизить периметром и площадью n-угольника.

Л е м м а 4. *Для любой замкнутой плоской кривой длины P^*, охватывающей площадь S^*, и для любого $\varepsilon > 0$ можно найти некоторый n-угольник, периметр P и площадь S которого удовлетворяют неравенствам*

$$|P - P^*| \leqslant \varepsilon, \quad |S - S^*| \leqslant \varepsilon. \tag{5}$$

Из леммы 4 и соотношения (3) получаем, что при любом ε найдется такой многоугольник с периметром P и площадью S, что выполнены неравенства

$$4\pi S^* \leqslant 4\pi S + 4\pi\varepsilon \leqslant P^2 + 4\pi\varepsilon \leqslant (P^* + \varepsilon)^2 + 4\pi\varepsilon = P^{*2} + \varepsilon(2P^* + 4\pi + \varepsilon).$$

В силу того, что ε произвольно, приходим к окончательному неравенству

$$4\pi S^* \leqslant P^{*2}, \tag{6}$$

которое согласно (4) превращается в равенство для круга.

Подытожим сказанное в виде следующего утверждения.

Т е о р е м а 2. *Площадь, охватываемая любой замкнутой кривой данной длины, не превосходит площади круга, окружность которого имеет ту же длину.*

Мы нашли полное решение изопериметрической проблемы.

К о м м е н т а р и и. 1. Полное решение нашей задачи получилось соединением двух геометрических лемм Зенодора и двух современных, но, по сути дела, технических лемм 3 и 4. Все необходимое для доказательства леммы 3 было заготовлено в трудах Вейерштрасса; понятиям длины кривой и площади, охватываемой кривой, было придано точное значение Жорданом, и тем самым им были сделаны основные заготовки для доказательства леммы 4.

2. Подробное проведение доказательств лемм 3 и 4 можно прочитать в цитировавшейся книге Бляшке [4].

Но прежде чем закончить этот рассказ — еще одно отступление.

Д о к а з а т е л ь с т в о Ш т е й н е р а. Трудно удержаться и наряду с доказательством, восходящим к идеям древних, не предложить схему еще одного доказательства, основная мысль которого принадлежит Якобу Штейнеру, математику, обогатившему геометрию многими замечательными идеями. Доказательство Штейнера неявно подразумевает, что искомая кривая, решающая изопериметрическую задачу, существует. (А мы ведь уже знаем, что это действительно так.) Остается доказать, что эта экстремальная кривая — окружность.

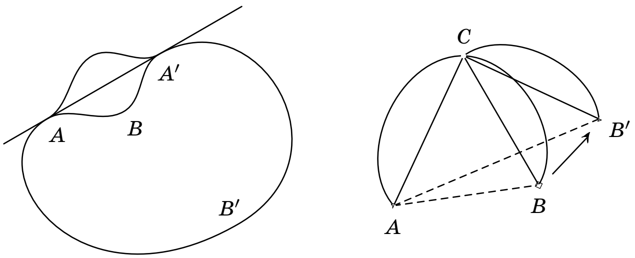
Утверждение 1. *Экстремальная кривая выпукла.*

Что такое выпуклая кривая? Это такая кривая, что если взять любые две точки, лежащие в области, ограниченной этой кривой, то и весь отрезок, соединяющий две точки, будет лежать внутри кривой.

Заметим кстати, что выпуклость играет большую роль в задачах на максимум и минимум. О ней у нас еще пойдет речь дальше. Выпуклости посвящено много замечательных книг, рассчитанных на школьников, например: Люстерник [14], Яглом и Болтянский [24] и др.

Вернемся к доказательству Штейнера и докажем утверждение 1.

Если кривая не выпукла, то на ней нашлись бы две точки A и A' такие, что обе дуги ABA' и $AB'A$, соединяющие точки A и A', лежат по одну сторону от прямой AA' (рис. 6). Заменив



одну из этих дуг ее зеркальным отражением относительно AA', получим новую кривую большей площади при той же длине.

Утверждение 2. *Если точки A и B делят длину экстремальной линии пополам, то хорда $[AB]$ делит площадь фигуры пополам.*

Действительно, если бы хорда $[AB]$ делила площадь на две неравные части, то бо́льшую часть следовало бы отразить относительно диаметра, и фигура, состоящая из большей части и ее отражения, имела бы ту же длину и большую площадь.

Утверждение 3. *Пусть снова точки A и B делят длину экстремальной линии пополам и C — любая точка кривой. Тогда угол ACB — прямой.*

Это центральное место. Метод, применяемый далее, носит название *четырехшарнирного метода Штейнера*.

Пусть имеется точка C такая, что угол ACB не является прямым. Площадь, ограниченная дугой ACB и диаметром AB, разбивается на три части: треугольник ABC и сегменты, прилегающие к сторонам AC и CB. Так вот, представим теперь себе, что в точке C у нас шарнир, соединяющий эти два сегмента. «Раздвинем» сегменты так, чтобы угол ACB' стал прямым (рис. 7). Тогда площадь, ограниченная дугой ACB' увеличится, ибо из всех треугольников с заданными боковыми сторонами наибольшую площадь имеет прямоугольный треугольник (так как $S_{\triangle ABC} = 0{,}5 \cdot |AC| \cdot |BC| \cdot \sin C \leqslant 0{,}5|AC| \cdot |BC|$ и равенство достигается, если угол равен 90°). А теперь отразим полученную фигуру относительно AB'. В итоге приходим к фигуре с тем же периметром и большей площадью. Утверждение доказано.

Мы пришли к следующему: *экстремальная кривая — это множество точек C, из которых хорда AB, делящая длину экстремальной линии пополам, видна под углом 90°, т.е. эта кривая — окружность.*

Человек восторженный воскликнет: «Потрясающе!» Скептик же начнет придираться: «То не доказано, это надо обосновать... Попробуй докажи существование... Почему при шарнирном раздвижении части сегментов у точки C не станут пересекаться?»... Не станем отвечать на его ворчанье. Признаем: потрясающе, но требует обоснования!

Классической изопериметрической задаче посвящена значительная литература: Курант и Роббинс [19], Крыжановский [12], Радемахер и Тёплиц [17].

Рассказ третий

МАКСИМУМЫ И МИНИМУМЫ В ПРИРОДЕ (ОПТИКА)

По Лейбницу наш мир является наилучшим из всех возможных миров, и потому его законы можно описать экстремальными принципами.

К. Зигель

Карл Зигель — выдающийся математик XX столетия. Ему принадлежат фундаментальные результаты во многих областях математики и механики. Слова, поставленные нами выше в качестве эпиграфа, — это, конечно, шутка, но в ней содержится доля истины. У нас уже был повод при обсуждении задачи Герона отметить, что природа «руководствуется» экстремальными принципами. В первом рассказе говорилось, что при отражении от плоской поверхности она «избирает» траекторию наименьшей длины.

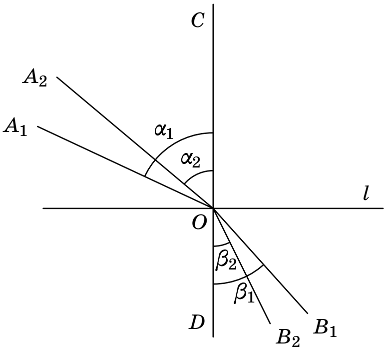
В словах Герона, процитированных нами в первом рассказе, можно увидеть зародыш фундаментального принципа, окончательно сформулированного в XVII—XIX веках. Тогда было уяснено, что природе свойственно «действовать» оптимально и в оптике, и в механике, и в термодинамике — вообще всюду.

Экстремальный принцип, касающийся явлений природы, был впервые четко сформулирован в оптике при попытке теоретического осмысления законов преломления света. Разнообразным вопросам оптики, в частности истории закона о преломлении света, посвящена книга Тарасова и Тарасовой [18].

Если опустить шест в неподвижную гладь прозрачного озера, он покажется нам как бы изломанным. Это происходит из-за преломления света.

Еще древние старались найти закон преломления. В частности, во II веке до н. э. Птолемей пытался найти этот закон опытным путем. Но он не дошел до правильного ответа.

Впервые его нашел голландский ученый Снеллиус. Сейчас имя Снеллиуса не так широко известно, как имена его великих современников — Декарта, Гюйгенса и Ферма, о которых нам дальше придется говорить. Самым известным фактом, связываемым с именем Снеллиуса, так и остался открытый им экспериментально (и не опубликованный при жизни) закон преломления света. Но в свое время Снеллиус был очень знаменит: для Кеплера



он был «гордостью геометров нашего века». Здесь, пожалуй, уместно напомнить, что в XVII веке математиков называли геометрами. Вернемся, однако, к преломлению света.

Закон Снеллиуса состоит в следующем.

Пусть два луча A_1OB_1 и A_2OB_2 (идущие «сверху вниз») преломляются в точке O (рис. 8). Углы α_1 и α_2, образованные прямыми A_1O и A_2O с вертикалью OC, называются, как мы уже знаем, *углами падения*. Углы β_1 и β_2, образованные прямыми B_1O и B_2O с вертикалью OD, называются *углами преломления*. Снеллиус установил, что

$$\frac{\sin\alpha_1}{\sin\beta_1} = \frac{\sin\alpha_2}{\sin\beta_2},$$

т.е. что *отношение синуса угла падения к синусу угла преломления есть величина постоянная, не зависящая от угла падения*.

К тому же самому закону независимо от Снеллиуса пришёл Декарт — один из величайших мыслителей и ученых Франции. У нас еще будет повод (в последнем рассказе этой части) порассуждать на тему о том, «ошибаются ли гении?». Так вот, Декарт был из тех гениев, кто «ошибался». Из его «ошибок», разбросанных по ниве науки, выросли впоследствии многочисленные животворные побеги.

Так и здесь, Декарт вывел закон преломления, опираясь на свою концепцию распространения световых лучей. Сама эта концепция не выдержала испытания временем, но из нее выкристаллизовался впоследствии закон сохранения количества движения.

Из теории Декарта следовало, что в более плотной среде, например в воде, скорость распространения света больше, чем в менее плотной, например в воздухе.

Этот факт многим показался сомнительным. Другое объяснение закона преломления, исходя из в точности противоположной посылки (в более плотной среде свет распространяется медленнее), дал Ферма. Имя Ферма знакомо всем из-за его «великой теоремы». Ферма и Декарт были соотечественниками и современниками. Они часто спорили друг с другом в поисках научной истины. Так было и здесь. В данном случае Ферма оказался прав — экспериментально было доказано, что в более плотной среде свет распространяется медленнее.

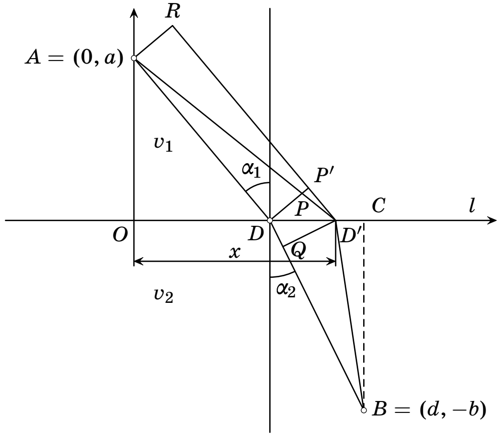
Для объяснения закона преломления света Ферма и выдвинул экстремальный принцип для оптических явлений. Впоследствии он был назван его именем. Принцип Ферма гласит: *в неоднородной среде свет избирает такую траекторию, вдоль которой время, затрачиваемое им на преодоление пути от одной точки до другой, минимально*.

Принцип Ферма позволяет точно поставить и решить задачу на минимум, приводящую к выводу закона Снеллиуса. А именно, этот принцип приводит к необходимости найти минимум функции одного переменного (рис. 9):

$$f(x) = \frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{b^2 + (d - x)^2}}{v_2}. \tag{1}$$

Надо сказать, что к моменту, когда Ферма выдвинул свой экстремальный принцип (а это произошло около 1660 г.), он уже владел алгоритмом нахождения максимумов и минимумов функций, состоящим, как теперь мы знаем, в приравнивании нулю производной. С помощью производных закон Снеллиуса выводится настолько просто, что его сейчас проходят даже в школе. Мы тоже повторим этот вывод в тринадцатом рассказе. Ферма же получил нужный результат гораздо более сложным путем. Может возникнуть вопрос: почему же он не воспользовался своим алгоритмом? Ответ очень прост: в то время производных еще не было! Лейбниц не опубликовал еще своей работы, где он ввел это понятие. Ферма мог применять свой прием только для полиномов, где он фактически предвосхитил понятие производной, а дифференцировать радикалы он не умел. И тот вывод закона Снеллиуса, который сейчас входит в школьный курс алгебры и анализа, был найден Лейбницем, причем в той же самой работе 1684 года, в которой заложен фундамент всего грандиозного строения — математического анализа.

Итак, Ферма вывел закон Снеллиуса из своего экстремального принципа, но его решение было весьма сложным. Гораздо более простое решение, основывающееся на принципе Ферма, дал Гюйгенс — еще один гениальный ученый XVII века, автор волновой теории света.



Решение Гюйгенса мы и приведем сейчас. Сначала необходимо точно поставить задачу. Она ставится так.

Даны две точки A и B по разные стороны от горизонтальной прямой l, разделяющей две среды. Требуется найти такую точку D, чтобы время преодоления пути ADB было минимальным при условии, что скорость распространения света в верхней среде v_1, а в нижней — v_2 (рис. 9). Отметим, что (1) есть математическая переформулировка задачи.

Обратим внимание на сходство этой задачи с задачей Герона.

Решение Гюйгенса. Пусть точка D (см. рис. 9) такова, что в ней выполнено соотношение

$$\frac{\sin\alpha_1}{\sin\alpha_2}=\frac{v_1}{v_2}. \tag{2}$$

Покажем, что для любой другой точки $D'\neq D$ время, затраченное на преодоление пути $AD'B$, будет больше времени, затраченного на путь ADB. Для этого восставим перпендикуляры в точках A и D к прямой AD. Точку пересечения с AD' перпендикуляра, проведенного из точки D, обозначим через P. Проведем прямую, параллельную AD, через точку D', и точку пересечения этой прямой и перпендикуляра DP обозначим P', а точку пересечения ее с перпендикуляром, проведенным из A, — через R. Наконец, опустим перпендикуляр $D'Q$ из D' на DB. Из рис. 9 видно, что величина угла PDD' равна α_1, а величина угла $D'DQ$ равна $\dfrac{\pi}{2}-\alpha_2$. Значит,

$$|D'P'|=|D'D|\sin\alpha_1,\quad |DQ|=|DD'|\sin\alpha_2. \tag{3}$$

Теперь проведем сравнение времен прохождения ломаных ADB и $AD'B$.

Вследствие того, что $|AP| > |AD|$, $|D'P| > |D'P'|$, $|D'B| > |BQ|$ (наклонные больше перпендикуляров), а также из (3) получаем

$$\frac{|AD'|}{v_1} > \frac{|AD| + |P'D'|}{v_1} = \frac{|AD|}{v_1} + |DD'|\frac{\sin\alpha_1}{v_1},$$
$$\frac{|D'B|}{v_2} > \frac{|BQ|}{v_2} = \frac{|DB| - |DQ|}{v_2} = \frac{|DB|}{v_2} - |DD'|\frac{\sin\alpha_2}{v_2}.$$

Складывая эти неравенства, получаем с учетом (2)

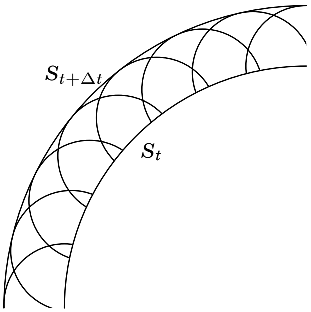
$$\frac{|AD'|}{v_1} + \frac{|D'B|}{v_2} > \frac{|AD|}{v_1} + \frac{|DB|}{v_2}.$$

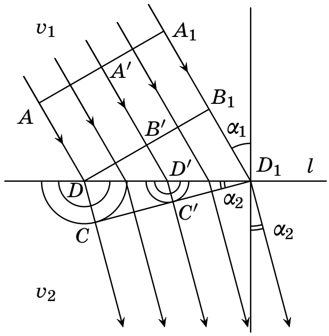
Итак, точка, преломляясь в которой свет потратит наименьшее время на прохождение пути от A к B, характеризуется тем, что отношение синуса угла падения к синусу угла отражения равно v_1/v_2, т. е. постоянному числу. Но именно в этом и состоит закон Снеллиуса.

В основу принципа Ферма положено допущение о том, что свет распространяется по некоторым линиям. Это представление легче всего увязать с корпускулярной теорией света, согласно которой свет — это *поток частиц*. Гюйгенсу принадлежит еще одно объяснение законов распространения и преломления света, основанное на представлении о свете как о *волне*, фронт которой движется со временем.

Волновой фронт S_t — это множество точек, которых свет, распространяемый некоторым источником, может достигнуть за заданное время t. Например, если в нулевой момент времени источник — это точка, а среда однородна, то через время t фронт S_t будет сферой радиуса vt с центром в источнике света. По мере удаления от источника сферическая волна становится все более плоской, и если мы вообразим источник бесконечно удаленным, то в пределе волновой фронт окажется плоскостью, равномерно движущейся со скоростью v.

Для определения движения волнового фронта в более сложных ситуациях Гюйгенс пользуется следующим правилом, получившим название «принцип Гюйгенса»: *каждая точка волнового фронта S_t сама становится вторичным источником, и через время Δt мы получаем семейство волновых фронтов от всех этих вторичных источников, а истинный волновой фронт $S_{t+\Delta t}$ в момент $t+\Delta t$ есть огибающая этого семейства* (т. е. поверхность, касающаяся всех вторичных волновых фронтов) (см. рис. 10).





Применим принцип Гюйгенса для вывода закона Снеллиуса.

Пусть параллельный пучок световых лучей падает на плоскую границу раздела двух однородных сред. Как и раньше, будем представлять себе, что l горизонтальна, а свет падает сверху (рис. 11). Через v_1 и v_2 (как и раньше) обозначим скорости распространения света над и под l и через α_1 и α_2 — углы падения и преломления соответственно. Волновой фронт $A_1A'A$ движется со скоростью v_1, и в некоторый момент t волновой фронт AA_1 достигает границы l в точке D. После этого D становится вторичным источником волн, распространяющихся в нижней среде со скоростью v_2. В точку D_1 свет придет в момент $t + |B_1D_1|/v_1 = t + (|DD_1|\sin\alpha_1)/v_1 = t_1$, а в промежуточную точку D' на отрезке DD_1 — в момент $t' = t + (|DD'|\sin\alpha_1)/v_1$. К моменту t_1 сферическая волна от вторичного источника D будет иметь радиус $r_1 = v_2(t_1 - t) = |DD_1|(v_2/v_1)\sin\alpha_1$, а волна от D' — радиус $r' = v_2(t_1 - t') = |D_1D'|(v_2/v_1)\sin\alpha_1$. Касательные D_1C и D_1C' к этим сферам совпадают, поскольку углы DD_1C и $D'D_1C$ равны (так как их синусы равны соответственно $r_1/|DD_1|$ и $r'/|D'D_1|$ и оба числа равны $(v_2/v_1)\sin\alpha_1$). Но точка D' была взята на DD_1 произвольно, и значит, вторичные волны в момент t_1 все касаются прямой CD_1, образующей с l угол α_2 такой, что $\sin\alpha_2 = (v_2/v_1)\sin\alpha_1$. Мы снова пришли к закону Снеллиуса.

Идея «волнового фронта» может быть проиллюстрирована не только на примере оптических задач. Вот другой случай. Пусть путник начинает свой путь из точки A, лежащей на прямолинейном шоссе, ограничивающем луг. На этом лугу имеется точка B, куда путник старается попасть возможно быстрее. Скорость по лугу v в два раза меньше скорости по шоссе. Если путник сразу пойдет по лугу, то за единицу времени он может попасть в любую точку окружности радиуса v. Если же он все это время будет идти по шоссе, то он пройдет расстояние $2v$. Пусть он часть пути пройдет по шоссе, а часть — по лугу. Тогда множество точек, где он сможет оказаться через единицу времени, образует «волновой фронт», состоящий из двух отрезков, соединенных дугой окружности. Отсюда очень легко ответить на вопрос, как оптимально двигаться путнику.

Здесь мне представляется уместным еще раз затронуть вопрос об экстремальных принципах.

В этом рассказе были даны два вывода закона преломления света. Между ними имеется принципиальное различие. При подходе Ферма никак не проясняется истинная сущность происходящего явления. Здесь постулируется некоторое свойство траекторий и показывается, что это допущение согласуется с экспериментом. Подход Гюйгенса отталкивается от описания физической природы явления.

Такая двойственность описания типична в естествознании. Законы природы, с одной стороны, допускают истолкования, базирующиеся на некоторых физических моделях, с другой — выводятся из экстремальных принципов.

Обсуждение причин завело бы нас слишком далеко, но в связи с этим хочу отметить, что описанные в нашем рассказе два подхода сыграли важнейшую роль в истории вариационного исчисления и всей теории экстремальных задач. На самом деле любая задача вариационного исчисления и оптимального управления может быть исследована двумя путями. Можно изучать ее экстремальные траектории (подобно Ферма), и это ведет к теории Эйлера—Лагранжа (которой мы коснемся в четырнадцатом рассказе). А кроме того, есть и другой путь (Гюйгенса): изучать пучки экстремальных траекторий, что приводит к аналогам волновых фронтов, к теории, разработанной Гамильтоном и Якоби в XIX столетии, и исследованию задач оптимального управления методами динамического программирования, которые в сравнительно недавние времена начал развивать американский ученый Беллман.

Рассказ четвертый

МАКСИМУМЫ И МИНИМУМЫ В ГЕОМЕТРИИ

История науки дает нам много примеров применения чистой геометрии и пользы, приносимой ею.

П. Лаплас

Архимеда будут помнить, когда Эсхила забудут, ибо языки умирают, а математические идеи — нет.

Г. Харди

Неиссякаемые россыпи драгоценных задач на максимум и минимум таятся в недрах древнейшей из математических наук — геометрии.

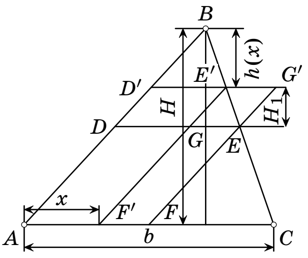
Геометрические задачи на максимум и минимум встречаются у всех трех величайших математиков античности — Евклида, Архимеда, Аполлония. Им воздавали дань крупнейшие математики эпохи Возрождения — Вивиани, Торричелли, Ферма и другие. Интерес к таким задачам сохранился до наших дней.

1. Задача Евклида. В «Началах» Евклида — первой научной монографии и первом учебном пособии в истории человечества, в труде, вышедшем в IV веке до н. э., имеется лишь одна задача на максимум. В современной редакции она выглядит так.

В данный треугольник ABC вписать параллелограмм $ADEF$ (*$EF \parallel AB$, $DE \parallel AC$*) *наибольшей площади* (рис. 12).

Приведем одно из возможных геометрических решений этой задачи, восходящее к решению Евклида, данному в «Началах».

Докажем, что *искомый параллелограмм характеризуется тем, что D, E и F — середины соответствующих сторон.*



Действительно, пусть $AD'E'F'$ — вписанный в ABC параллелограмм, отличный от $ADEF$. Точку пересечения прямых $D'E'$ и EF обозначим через G', а точку пересечения прямых DE и $E'F'$ — через G.

Покажем, что площадь параллелограмма $AD'E'F'$ меньше площади параллелограмма $ADEF$ на величину площади параллелограмма $EG'E'G$. Для этого проведем в треугольнике ABC из точки B высоту, длину которой обозначим через H. Длину стороны AC обозначим через b, а длину высоты треугольника $GE'E$, проведенной из точки E', — через H_1.

Из подобия треугольников GEE' и ABC ($E'G \parallel AB$, $GE \parallel AC$) получаем

$$\frac{H_1}{|GE|} = \frac{H}{b} \iff \frac{H_1}{H/2} = \frac{|GE|}{b/2}.$$

Из полученного соотношения следует, что площадь параллелограмма $D'G'ED$, высота которого H_1, а длина стороны DE — $b/2$, равна площади параллелограмма $EGF'F$, ибо его высота равна $H/2$, а длина стороны $F'F$ равна $|GE|$. Отсюда и следует, что площадь параллелограмма $ADEF$ равна площади фигуры $AD'G'EGF'$, т. е. на величину площади параллелограмма $GE'G'E$ больше, чем площадь $AD'E'F'$. Задача решена.

2. Задача Архимеда. Мы упоминали уже, что в некоторых сочинениях древних авторов говорится, что Архимед (287—212 гг. до н. э.) доказывал изопериметрическое свойство окружности и изопифанное свойство сферы. Однако в дошедших до нас сочинениях Архимеда изопериметрическая задача не упоминается, и вклад Архимеда в решение этой проблемы доныне неизвестен. Решение же одной *изопифанной* задачи имеется в сочинении Архимеда «О шаре и цилиндре». Там ставится и решается следующая проблема.

Найти шаровой сегмент, вмещающий максимальный объем среди всех сегментов, имеющих заданную площадь сферической поверхности.

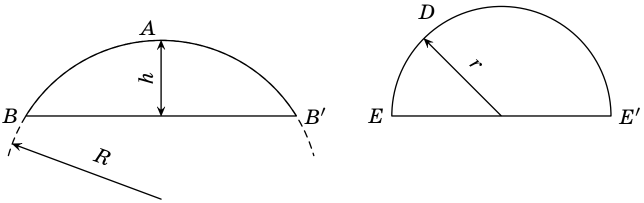
Приведем сначала решение, хотя целиком и основанное на идеях Архимеда, но все-таки сильно алгебраизированное. А потом мы приведем это же решение, изложенное на чисто геометрическом языке его автора.

Рассмотрим шар радиуса R и его шаровой сегмент BAB' высоты h (рис. 13). Наряду с сегментом BAB' рассмотрим полушар EDE' той же боковой поверхности. Обозначим его радиус через r. Объем V шарового сегмента равен, как мы знаем, $\pi h^2(R - h/3)$, площадь его боковой поверхности равна $2\pi Rh$, объем $\widehat{V}$ полушара равен $(2/3)\pi r^3$, площадь $\widehat{S}$ боковой поверхности — $2\pi r^2$. Из равенства боковых поверхностей сегмента и полушара получаем

$$r^2 = Rh. \qquad (1)$$

Докажем неравенство

$$(2R - r)r > (2R - h)h \quad \text{при} \quad h \neq R. \qquad (2)$$



Рассмотрим два случая: а) $h < R$ и б) $h > R$. В случае а)

$$r^2 = Rh > h^2 \implies r > h \implies R - r < R - h \implies (2R - r)r = \\ = R^2 - (R - r)^2 > R^2 - (R - h)^2 = (2R - h)h;$$

в случае б)

$$\left.\begin{matrix} r^2 = Rh < h^2 \\ r^2 = Rh > R^2 \end{matrix}\right\} \implies R < r < h \implies r - R < h - R \implies (2R - r)r = \\ = R^2 - (R - r)^2 > R^2 - (R - h)^2 = (2R - h)h.$$

Складывая (1) и (2) и умножая на $\pi h/3$, получаем

$$\frac{\pi h}{3} \cdot 2Rr > \frac{\pi}{3}(3R - h)h^2. \qquad (3)$$

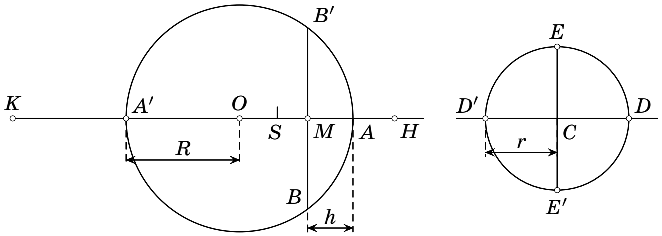
Заменив теперь в (3) Rh на r^2, приходим к нужному неравенству:

$$\widehat{V} = \frac{2}{3}\pi r^3 = \frac{\pi h}{3} 2Rr > \pi h^2 \left(R - \frac{h}{3}\right) = V.$$

Итак, полушар той же боковой поверхности имеет больший объём в сравнении с шаровым сегментом, или, говоря словами Архимеда, «из всех сферических сегментов, ограниченных равными поверхностями, наибольшим будет полушарие» (Архимед. Сочинения. — М.: Физматлит, 1962. С. 95—117).

Среди ученых всех времен гений Архимеда, наряду с гением Ньютона, вызывает, пожалуй, наибольшее восхищение. Не пожалеем же места и еще раз проведем решение нашей задачи, уже почти буквально следуя за мыслью Архимеда (в скобках мы все-таки указываем соответствующие алгебраические соотношения).

Архимед не мог пользоваться алгебраическим языком и алгебраическими выкладками — ведь до зарождения алгебры оставалось восемнадцать веков. Язык Архимеда — язык геометрии. На прямой $A'A$ (рис. 14) отложим, следуя Архимеду, отрезок $[OH]$ такой величины, чтобы конус с высотой HM и с радиусом основания MB был бы равновелик шаровому сегменту BAB'. На продолжении отрезка $[OA']$ отложим отрезок $[A'K]$, равный по



длине радиусу R. Равновеликость конуса и сегмента приводит Архимеда к следующей пропорции:

$$\frac{|HM|}{|AM|} = \frac{|KM|}{|A'M|}. \qquad (4)$$

Проверим, что это равенство действительно имеет место (используя известные формулы объема конуса $V_{\text{к}}$ и сегмента $V_{\text{с}}$):

$$V_{\text{к}} = \frac{\pi}{3}|HM||MB|^2 = \frac{\pi}{3}|HM||MA'||MA| = V_{\text{с}} =$$
$$= \frac{\pi}{3}(3R-h)h^2 = \frac{\pi}{3}|KM||AM|^2. \qquad (5)$$

Мы воспользовались тем, что длина отрезка $[MB]$ есть среднее геометрическое длин отрезков $[A'M]$ и $[MA]$. Из (5) формула (4) следует сразу.

Равенство поверхностей полушара и сегмента приводит к тому, что

$$|AB| = |ED|. \qquad (6)$$

Действительно, $|ED| = r\sqrt{2}$, $|AB|^2 = |AA'||AM|$ (снова — известное свойство треугольника, вписанного в окружность и опирающегося на диаметр), значит,

$$\pi|AB|^2 = 2\pi Rh = S_{\text{с}} = \widehat{S} = 2\pi r^2 = \pi|ED|^2 \implies |AB| = |ED|.$$

Далее Архимед откладывает отрезок $[AS]$, равный по длине $[CD]$ и доказывает неравенство (2):

$$|A'S||AS| > |A'M||AM| \quad (\Longleftrightarrow (2R-r)r > (2R-h)h).$$

Этот факт Архимед обосновывает геометрически: из двух прямоугольников с одинаковым периметром площадь больше у того, у которого больше длина меньшей стороны.

Далее имеем (из-за равенства боковых поверхностей сегмента и полушара)

$$|AS|^2 = |AM||A'K| \quad (\Longleftrightarrow r^2 = Rh).$$

Складывая последнее неравенство с этим равенством, получаем

$$|AS||AA'| > |KM||AM| \quad (\Longleftrightarrow 2Rr > (3R-h)h).$$

Умножая на $|AM|$ и используя (5), получим

$$|AS||AA'||AM| > |KM||AM|^2 \quad (\Longleftrightarrow 2Rrh > (3R-h)h^2). \qquad (7)$$

Ранее у нас были доказаны соотношения

$$|KM||AM|^2 = |HM||MB|^2 \quad (\text{см. } (5)),$$
$$|AA'||AM| = |AB|^2 = |ED|^2 \quad (\text{см. } (6)),$$

$|AS| = |CD|$ (по построению), откуда и из (7) получим

$$\widehat{V} = \frac{\pi}{3}|CD||ED|^2 > \frac{\pi}{3}|HM||MB|^2 = V_{\text{к}} = V_{\text{c}}$$
$$\left(\Longleftrightarrow \widehat{V} = \frac{2\pi}{3}r^3 > \frac{\pi}{3}(3R-h)h^2 = V_{\text{c}}\right).$$

Доказательство завершено.

Может быть, нелишним будет напомнить, что все формулы, которые были использованы нами (объем конуса, шара и шарового сегмента, площадь поверхности шара и сферической поверхности сегмента), были также впервые получены именно Архимедом и все в той же работе «О шаре и цилиндре». Трудно не согласиться с Харди — Архимед будет славен, доколь жив будет хоть один математик. Со второй же частью его фразы мне не хочется соглашаться — да будет славен и Эсхил.

Разговор о задаче, составленной и решенной Аполлонием, отложим до тринадцатого рассказа.

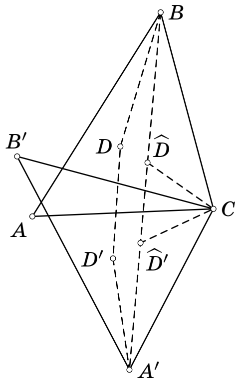
3. З а д а ч а Ш т е й н е р а. *В плоскости треугольника найти точку, сумма расстояний от которой до вершин треугольника минимальна.*

В иной формулировке эта задача обсуждалась в первом рассказе. Она тоже имеет давнюю историю, хотя и не такую, как героновская и классическая изопериметрическая. Она была помещена в сочинении Вивиани «О максимальных и минимальных значениях». Это был первый труд, специально посвященный нашему предмету. Он вышел в свет в 1659 году.

Этой же задачей интересовались Кавальери и Торричелли. (Само же решение в этой задаче, т. е. точку, где достигается искомый минимум, принято называть *точкой Торричелли* — см., например, книгу Зетеля [9]); Кокстер [10] утверждает, что нашей задачей занимался и Ферма. (Вивиани, Кавальери и Торричелли — крупнейшие итальянские ученые XVII века. Кавальери известен своим принципом — предтечей интегрального исчисления, Торричелли — открытием атмосферного давления. Торричелли и Вивиани — ученики Галилея. Именно Вивиани ослепший Галилей диктовал в конце жизни свои «Беседы о механике».)

Интерес столь крупных ученых к такой элементарной задаче — еще одно подтверждение того, что стимулом к творчеству нередко являются эстетические мотивы.

В XIX столетии этой и рядом подобных проблем много занимался Штейнер. Их часто именуют *проблемами Штейнера*. Так поступим и мы.

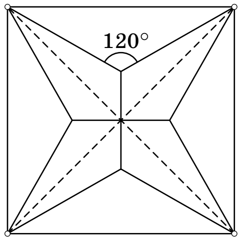


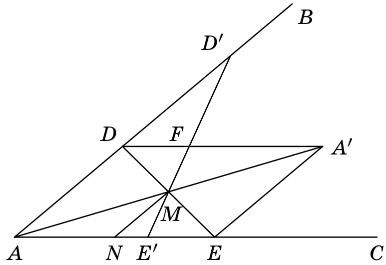
Приведем известное геометрическое решение задачи Штейнера для треугольника с углами, не превосходящими 120°.

Пусть в треугольнике ABC (рис. 15) угол $C \geqslant 60^\circ$. Повернем теперь треугольник ABC вокруг точки C на угол 60°. Получим треугольник $A'B'C$. Возьмем любую точку D в треугольнике ABC, а через D' обозначим образ D при нашем повороте. Тогда сумма длин $|AD| + |BD| + |CD|$ равна длине ломаной $|BD| + |DD'| + |D'A'|$.

Пусть теперь $\widehat{D}$ — точка Торричелли, т. е. точка, из которой все стороны треугольника видны под углом 120°, и $\widehat{D}'$ — образ $\widehat{D}$ при повороте. Нетрудно понять, что точки B, $\widehat{D}$, $\widehat{D}'$ и A' лежат на одной прямой, откуда следует, что точка Торричелли и является решением задачи. Отдельно нужно еще доказать, что если тупой угол больше 120°, то решением задачи будет вершина тупого угла. Сделайте это самостоятельно.

Две задачи (четвертая и пятая) из первого рассказа примыкают к штейнеровской. Задачу 4 из первого рассказа продумайте самостоятельно. Ответ в ней таков: если точки A, B, C, D образуют выпуклый четырехугольник, то искомая точка — точка пересечения диагоналей, если невыпуклый, то это вершина наибольшего угла. Что же касается задачи 5, то ответ там отрицательный. К примеру, для квадрата имеются две экстремальные сети, изображенные нами на рис. 16. Сумма длин этих сетей меньше, чем сумма длин диагоналей.





Приведем еще две известные геометрические задачи на минимум.

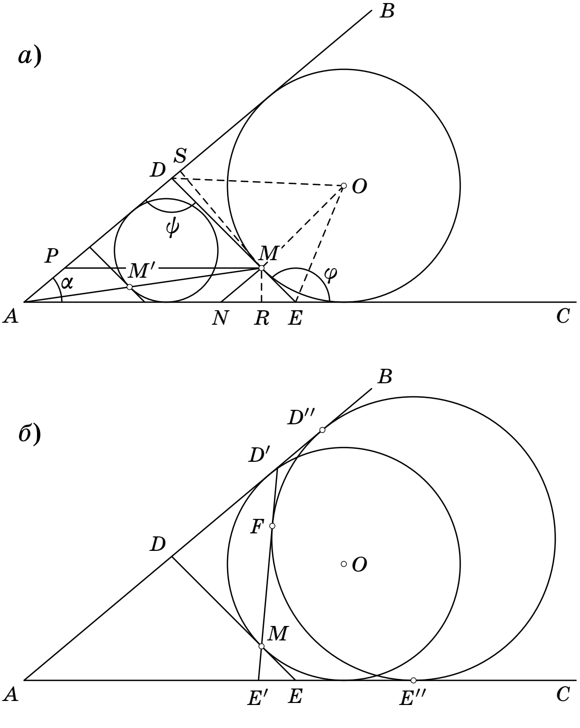
4. З а д а ч а о н а и м е н ь ш е й п л о щ а д и. *Дан угол и точка внутри него. Требуется провести через эту точку прямую, отсекающую от угла треугольник наименьшей площади.*

Покажем, что искомая прямая обладает тем свойством, что отрезок ее, лежащий внутри угла, делится заданной точкой пополам. Такую прямую нетрудно построить. Можно, например, соединить заданную точку M (см. рис. 17) с вершиной A, на продолжении отрезка $[AM]$ отложить отрезок $[MA']$, равный по длине отрезку $[AM]$, и провести через точку A' прямую параллельно AC. Пусть D — точка пересечения этой прямой и стороны AB. Тогда, как легко понять, прямая, соединяющая точку D с точкой M и пересекающая AC в точке E, обладает тем свойством, что $|DM| = |ME|$ (ибо треугольники MDA' и MEA равны). Искомая прямая построена. Возможны и другие способы построения.

Докажем теперь, что построенная прямая действительно является искомой. Для этого проведем какую-нибудь еще прямую $D'E'$. Пусть для определенности точка E' лежит левее E. Тогда площадь треугольника $AE'D'$ равна площади треугольника AED минус площадь треугольника EME' и плюс площадь треугольника MDD'. Обозначим через F точку пересечения прямой D' с прямой $D'E'$. Тогда треугольники EME' и MDF равны. Но второй из этих треугольников содержится в треугольнике $DD'M$. Из сказанного вытекает, что площадь треугольника ADE меньше площади треугольника $AD'E'$.

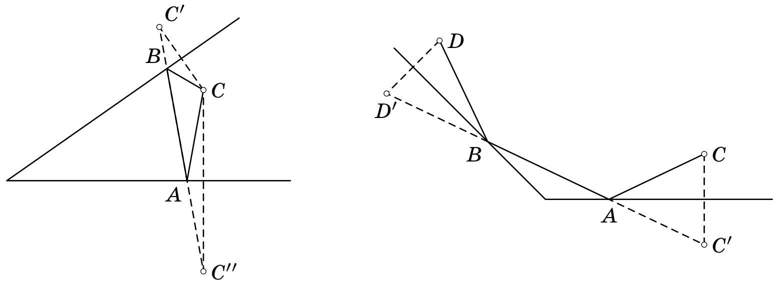
5. З а д а ч а о н а и м е н ь ш е м п е р и м е т р е. *Дан угол и точка M внутри него. Требуется провести через эту точку прямую, отсекающую от угла треугольник наименьшего периметра.*

Покажем, что искомая прямая DE обладает тем свойством, что вневписанная в треугольник ADE окружность касается отрезка $[DE]$ в точке M. Такую прямую нетрудно построить. Впишем в угол BAC некоторую окружность (рис. 18, *а*) и обозначим через M' ближайшую к A точку пересечения прямой AM с этой окружностью. Далее следует провести касательную к построенной окружности в точке M' и затем — прямую через M, параллельную этой касательной. Эта прямая и будет искомой.



А теперь докажем, что построенная прямая действительно дает решение. Проведем еще какую-нибудь прямую $D'E'$ (рис. 18, *б*). Пусть вневписанная в треугольник $AD'E'$ окружность касается отрезка $D'E'$ в точке F, а сторон угла AB и AC — в точках D'' и E'' соответственно. Длины отрезков $[E'F]$ и $[E'E'']$ равны как длины отрезков касательных, проведенных из одной точки. То же самое можно сказать о длинах отрезков $[D'F]$ и $[D'D'']$. Значит, периметр треугольника $AD'E'$ равен сумме длин отрезков $[AD'']$ и $[AE'']$, при этом, конечно, $|AD''| = |AE''|$. Таким образом, периметр треугольника $AD'E'$ будет наименьшим, когда точки E'' и D'' будут возможно близко к A. Но это произойдет как раз тогда, когда вневписанная окружность «упрется» в точку M, т. е. когда отрезок прямой, проходящей через точку M, будет касаться этой окружности в точке M.

У нас остался еще один геометрический долг: не обсуждены задачи 1 и 2 из первого рассказа. Решения этих задач ясны из рис. 19 и 20. Если же построения, изображенные на этих рисунках, невозможны, то искомые точки должны совпадать с вершиной угла.



Нашу тему исчерпать невозможно. Ей посвящено огромное количество книг. Назовем некоторые из них. Там читатель найдет любую задачу себе по вкусу. А потом, после того как он прочтет вторую часть, он сможет попытаться найти аналитические решения. Вот список книг, где имеются экстремальные задачи геометрии: Зетель [9], Кокстер [10], Курант и Роббинс [13], Шарыгин [21], [22], Шклярский, Ченцов, Яглом [23]. Специально этим вопросам посвящена статья Болтянского и Яглома [6] в «Энциклопедии элементарной математики». В тринадцатом рассказе будет разобрано еще несколько геометрических задач.

Рассказ пятый

МАКСИМУМЫ И МИНИМУМЫ В АЛГЕБРЕ И АНАЛИЗЕ

Алгебра щедра. Зачастую она дает больше, чем у нее спрашивают.

Ж. Даламбер

1. Задача Тартальи. Начнем наш рассказ с обсуждения следующего вопроса, который был поставлен в одном из сочинений Никколо Тартальи.

Разделить число восемь на две такие части, чтобы произведение их произведения на их разность было максимальным.

Тарталья (1500—1557) вошел в историю науки как человек, научившийся решать уравнения третьей степени.

Сделаем попытку восстановить ход рассуждений Тартальи при решении его задачи. Для этого нелишне будет сказать несколько слов об истории его замечательного открытия.

Решение уравнения $x^3 + px + q = 0$ (для положительного p и отрицательного q) нашел впервые итальянский математик Сципион дель Ферро, живший на рубеже XV и XVI веков. Это было время, когда корнями уравнения признавались лишь положительные корни — ни отрицательные, ни тем более комплексные в расчет не брались. Ферро не опубликовал своего открытия, но посвятил в него нескольких близких себе людей. В те времена были очень распространены «математические бои». (В наши дни эта традиция возрождена — правда, в несколько иной форме. Вместо единоборств отдельных личностей в бой вступают целые коллективы: интернаты друг с другом, победители школьной олимпиады и ее жюри и т. п.)

Один из посвященных в тайну решения уравнения третьей степени решил воспользоваться ею, чтобы одерживать верх в математических сражениях. И он, наверное, добился бы полного успеха, если бы однажды судьба не свела его с Никколо Тартальей. Тарталье было предложено решить 30 уравнений третьей степени при различных значениях p и q. Сначала Тарталья не знал, что его противник обладает тайной общего решения. Но незадолго до окончания того срока, когда необходимо было представить решения задач, Тарталья узнал об этом. Вложив все свои силы в решение проблемы, Тарталья за 8 дней до назначенного срока самостоятельно нашел этот способ. (Много подробностей обо всем этом читатель узнает из интересной книги Гиндикина [7].) Иначе говоря, он получил (вслед за Ферро) следующую формулу:

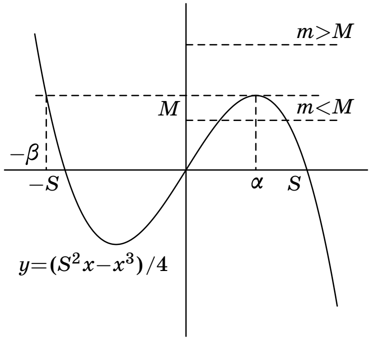
$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{1}$$

Формула (1) дает выражение для положительного корня в случае Ферро (если $p > 0$, а $q < 0$). Но и в других случаях (например, в том, который нам встретится дальше, когда $p < 0$, а $q > 0$) она также дает выражение для вещественного корня. Эта формула называется обычно *формулой Кардано*, по имени человека, впервые опубликовавшего ее. Тарталья самой формулы не опубликовал, но в ряде своих работ он сообщал о том, что умеет решать те или иные уравнения и задачи. В одном из его сочинений была поставлена и наша задача. Автор не привел ее решения, но указал ответ. Этот ответ был сформулирован так: «*число* 8 *следует разделить пополам; квадрат этой половины, увеличенный на треть этого квадрата, должен равняться квадрату разности обеих частей*». Таким образом, если искомые числа обозначить через a и b ($a > b$), то для разности $a - b$ Тарталья дал такое выражение: $(a-b)^2 = (8:2)^2 + (8:2)^2 : 3 = 64/3$, т. е. $a - b = 8/\sqrt{3} \implies a = 4 + 4/\sqrt{3}$, $b = 4 - 4\sqrt{3}$. Как мы увидим дальше, Тарталья не ошибся.

А теперь попробуем (следуя книге Цейтена [20]) восстановить ход мыслей Тартальи, приведший его к правильному ответу. Не будем связывать себя с конкретным числом 8 и будем решать задачу в общем виде. Пусть требуется разделить произвольное число S. Как мы видели, Тарталья в ответе указывает не самые числа a и b, а их р а з н о с т ь. Наверное, именно ее он принимал за неизвестное. Итак, пусть $a - b = x$, тогда $a = (S + x)/2$, $b = (S - x)/2$, и значит, требуется найти максимум функции $f(x) = x(S/2 + x/2)(S/2 - x/2) = (S^2x - x^3)/4$. Обозначим максимальное значение этой функции (для $x \geqslant 0$) через M. Мы пришли к уравнению для x:

$$x\left(\frac{S}{2} + \frac{x}{2}\right)\left(\frac{S}{2} - \frac{x}{2}\right) = M \iff x^3 - S^2x + 4M = 0. \tag{2}$$

К сожалению, уравнение (2) не имеет структуры уравнения Ферро, ибо здесь $p = -S^2 < 0$, а $q = 4M > 0$. Но, с другой стороны, уравнение (2) имеет примечательную особенность: помимо



отрицательного корня (обозначим его $-\beta$) оно имеет *двукратный положительный корень*, т. е. здесь обращаются в нуль функция и производная. Из рис. 21 видно, что при $m > M$ уравнение $x^3 - S^2x + 4m = 0$ не имеет положительных корней, при $m < M$ имеет два корня, а при $m = M$ — один положительный корень. Положительный корень уравнения (2) обозначим через α — это число и будет равно искомой разности. Таким образом, мы можем записать тождество

$$x^3 - S^2x + 4M = (x+\beta)(x-\alpha)^2 = x^3 + (\beta - 2\alpha)x^2 + (\alpha^2 - 2\alpha\beta)x + \alpha^2\beta,$$

из которого следует, что

$$\beta = 2\alpha,\ \ p = -S^2 = \alpha^2 - 2\alpha\beta = \alpha^2 - 4\alpha^2 = -3\alpha^2,\ \ q = 4M = \alpha^2\beta = 2\alpha^3,$$

и значит,

$$\frac{q^2}{4} + \frac{p^3}{27} = 0 \iff \frac{(4M)^2}{4} = \frac{S^6}{27} \iff (2M)^2 = \left(\frac{S^2}{3}\right)^3.$$

При этом формула (1) дает выражение для отрицательного корня:

$$-\beta = 2\sqrt[3]{-\frac{q}{2}} = -2\sqrt[3]{2M} = -2\frac{S}{\sqrt{3}} \iff \beta = \frac{2S}{\sqrt{3}}.$$

Отсюда

$$\alpha = \frac{\beta}{2} = \frac{S}{\sqrt{3}} \implies \alpha^2 = \frac{S^2}{3} = \left(\frac{S}{2}\right)^2 + \frac{1}{3}\left(\frac{S}{2}\right)^2. \qquad (3)$$

Если $S = 8$, то и выходит, что для получения квадрата разности «число 8 следует разделить пополам и квадрат этой половины увеличить на одну треть этого квадрата»... Задача оказалась решенной.

Много интересных задач на максимум и минимум скрыто в различного рода «точных неравенствах». Продолжим наш рассказ обсуждением, пожалуй, самого древнего такого неравенства.

2. Неравенство между средним геометрическим и средним арифметическим двух чисел. Пусть a и b — неотрицательные числа. Их *средним геометрическим* называется число $\sqrt{ab}$, а *средним арифметическим* — число $(a+b)/2$. Докажем, что

для любых неотрицательных чисел a и b имеет место неравенство

$$\sqrt{ab} \leqslant \frac{a+b}{2}, \tag{1}$$

т. е. *что среднее геометрическое не превосходит среднего арифметического.* Неравенство (1) *точное*. Это означает, что в (1) иногда достигается равенство, а именно, это бывает тогда (и только тогда), когда $a = b$.

В неравенстве (1) на самом деле таятся разные экстремальные задачи. Приведем две.

А) *Найти максимум произведения двух чисел, если их сумма постоянна.*

Б) *Найти максимальную площадь прямоугольного треугольника, если сумма длин катетов постоянна.*

Из неравенства (1), в частности, вытекает, что *среди прямоугольных треугольников с заданной суммой катетов максимальную площадь имеет равнобедренный треугольник*.

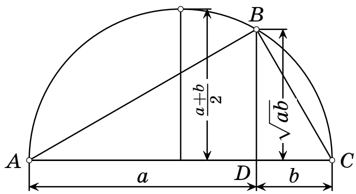
Задача А является алгебраической по своему содержанию; задача Б — геометрической. Когда Ферма открыл свой метод нахождения максимумов и минимумов (о нем мы побеседуем в одиннадцатом рассказе), он изложил его в частном письме известному математику того времени Робервалю. Иллюстрацию своего метода он привел, решив задачу Б, сам же геометрический факт был известен еще античным геометрам.

Неравенство (1) можно доказать разными способами. Приведем здесь два доказательства: алгебраическое и геометрическое.

Алгебраическое решение основано на следующей цепочке очевидных неравенств:

$$0 \leqslant (a-b)^2 \implies 2ab \leqslant a^2+b^2 \implies \\ \implies 4ab \leqslant a^2+2ab+b^2 = (a+b)^2 \implies \sqrt{ab} \leqslant \frac{a+b}{2}.$$

Предоставим слово геометрии (рис. 22). Возьмем отрезок $[AC]$ длины $a+b$ ($|AD|=a$, $|DC|=b$) и проведем полуокружность, опирающуюся на этот отрезок. Из точки D восставим перпендикуляр к AC. Пусть B — точка пересечения этого перпендикуляра с полуокружностью. Из подобия треугольников ABD и BCD (надо иметь в виду, что угол B опирается на полуокружность и, следовательно, прямой, т. е. угол A равен углу DBC, а угол C равен углу ABD) получаем



$$\frac{|AD|}{|BD|}=\frac{|BD|}{|DC|} \implies |BD|=\sqrt{ab}.$$

Если теперь при фиксированном отрезке $[AC]$ (т. е. при заданной сумме $a+b$) менять точку D, то видно, что отрезок BD будет иметь максимальную длину (а именно равную $(a+b)/2$) тогда и только тогда, когда точка D совпадает с центром полуокружности. Этим и доказывается неравенство (1).

3. Неравенство между средним геометрическим и средним арифметическим (общий случай). Докажем следующую теорему.

Для любых неотрицательных чисел $x_1, \ldots, x_n$ *имеет место неравенство*

$$\sqrt[n]{x_1 \ldots x_n} \leqslant \frac{x_1+\ldots+x_n}{n}. \tag{1}$$

Левая часть неравенства (1) называется *средним геометрическим чисел* $x_1, \ldots, x_n$, правая называется их *средним арифметическим*. Таким образом, не только для $n=2$, но и при любом n *среднее геометрическое не превосходит среднего арифметического*. Неравенство (1) является *точным*. Если все числа равны, оно обращается в равенство.

Существует большое число доказательств неравенства (1). Одно из самых красивых и совершенно элементарное — доказательство, данное знаменитым французским математиком О. Коши.

Сначала покажем, как по методу Коши доказать неравенство (1) для $n=3$. Для этого выведем (1) при $n=4$, а затем «спустимся» к $n=3$. Для $n=4$ неравенство (1) сразу следует из двукратного применения доказанного в предыдущем пункте неравенства (1) (для $n = 2$):

$$x_1 \cdot x_2 \cdot x_3 \cdot x_4 = (x_1 \cdot x_2) \cdot (x_3 \cdot x_4) \leqslant \left(\frac{x_1 + x_2}{2}\right)^2 \left(\frac{x_3 + x_4}{2}\right)^2 =$$

$$= \left[\left(\frac{x_1 + x_2}{2}\right)\left(\frac{x_3 + x_4}{2}\right)\right]^2 \leqslant \left(\frac{x_1 + x_2 + x_3 + x_4}{4}\right)^4, \quad (2)$$

т. е. неравенство (1) для $n = 4$ доказано.

А теперь получаем (применив (2))

$$(x_1 \cdot x_2 \cdot x_3)^{1/3} = [x_1 \cdot x_2 \cdot x_3 \cdot (x_1 \cdot x_2 \cdot x_3)^{1/3}]^{1/4} \leqslant$$

$$\leqslant \frac{x_1 + x_2 + x_3 + (x_1 x_2 x_3)^{1/3}}{4} \implies \frac{3}{4}(x_1 \cdot x_2 \cdot x_3)^{1/3} \leqslant$$

$$\leqslant \frac{x_1 + x_2 + x_3}{4} \implies (x_1 \cdot x_2 \cdot x_3)^{1/3} \leqslant \frac{x_1 + x_2 + x_3}{3},$$

что доказывает (1) для $n = 3$.

Теперь докажем (1) в общем случае. Во-первых, отметим что точно так же, как выше было доказано неравенство (1) для $n = 4$, можно доказать его для $n = 8$, затем для $n = 16$ и т. д. — для любого $n = 2^k$, $k = 2, 3, \ldots$

Применим «*метод спуска*», который был уже использован нами при переходе от четырех к трем. Пусть неравенство для $n = m+1$ уже доказано. Докажем его для $n = m$. Имеем по нашему допущению

$$(x_1 \ldots x_m)^{1/m} = [(x_1 \ldots x_m)(x_1 \ldots x_m)^{1/m}]^{1/(m+1)} \leqslant$$

$$\leqslant \frac{x_1 + \ldots + x_m + (x_1 \ldots x_m)^{1/m}}{m+1}.$$

Отсюда

$$(x_1 \ldots x_m)^{1/m}\left(1 - \frac{1}{m+1}\right) \leqslant \frac{x_1 + \ldots + x_m}{m+1} \implies$$

$$\implies (x_1 \ldots x_m)^{1/m} \leqslant \frac{x_1 + \ldots + x_m}{m}.$$

Неравенство (1) полностью доказано.

Это доказательство — одно из очень многих. В известной книге Беккенбаха и Беллмана «Неравенства» (М.: ИЛ, 1948) приведено двенадцать доказательств этого неравенства. Пожалуй, самое простое из них принадлежит Элерсу. Докажем по индукции, что из $x_1 \ldots x_n = 1$, $x_i > 0$, следует неравенство $x_1 + \ldots + x_n \geqslant n$ (отсюда все получается очевидным образом). Для $n = 1$ это утверждение тривиально. Пусть для $n = m$ утверждение доказано, и допустим, что $x_1 \ldots x_{m+1} = 1$. Тогда имеются два числа (пусть это x_1 и x_2) такие, что $x_1 \geqslant 1$, а $x_2 \leqslant 1$, т. е. $(x_1 - 1)(x_2 - 1) \leqslant 0$, или $x_1x_2+1 \leqslant x_1+x_2$. Отсюда и из допущения индукции вытекает $x_1+\ldots+x_{m+1} \geqslant 1+x_1x_2+x_3+\ldots+x_{m+1} \geqslant 1+m$, что и требовалось доказать.

Неравенство между средним арифметическим и средним геометрическим издавна было любимой темой кружковых занятий. Например, рассмотрим следующую задачу.

В данный шар вписать конус наибольшего объема.

Обозначим через R радиус шара и через r и h — радиус основания и высоту конуса соответственно. Тогда (продумайте это) объем V конуса равен $\pi h^2(2R-h)/3$. Применяя неравенство между средним арифметическим и средним геометрическим, получим

$$\frac{3V}{4\pi} = \frac{h}{2}\frac{h}{2}(2R-h) \leqslant (2R)^3,$$

и равенство достигается при $h/2 = 2R-h \implies h = (4/3)R$. При этой высоте объем конуса и будет максимальным.

Вот еще две задачи.

В данный конус вписать цилиндр максимального объема.

Дан прямоугольный лист жести размерами $a \times b$. Требуется вырезать около всех его углов одинаковые квадратики так, чтобы после загибания остающихся кромок получилась открытая сверху коробка наибольшей вместимости.

Впрочем, решать такие задачи нашим приемом интересно лишь до того, как научишься дифференцировать.

4. Неравенство между средним арифметическим и средним квадратическим. Пусть $x_1, \ldots, x_n$ — некоторые числа. Их *средним квадратическим* называется число $[(x_1^2+\ldots$ $\ldots+x_n^2)/n]^{1/2}$. Имеет место следующая теорема.

Для любых чисел $x_1, \ldots, x_n$ выполняется неравенство

$$\frac{x_1+\ldots+x_n}{n} \leqslant \left(\frac{x_1^2+\ldots+x_n^2}{n}\right)^{1/2}, \tag{1}$$

т. е. *среднее арифметическое всегда не превосходит среднего квадратического.* Неравенство (1) является *точным*. Если все числа равны, оно обращается в равенство.

Неравенство (1) можно также доказывать по-разному. Но самое простое, пожалуй, такое доказательство. Имеем

$$0 \leqslant (a-b)^2 \implies 2ab \leqslant a^2+b^2. \tag{2}$$

Возведя среднее арифметическое в квадрат и воспользовавшись далее неравенством (2), получим

$$\left(\frac{x_1+\ldots+x_n}{n}\right)^2=\frac{x_1^2+\ldots+x_n^2+2x_1x_2+2x_1x_3+\ldots+2x_{n-1}x_n}{n^2}\leqslant$$
$$\leqslant\frac{x_1^2+\ldots+x_n^2+(x_1^2+x_2^2)+(x_1^2+x_3^2)+\ldots+(x_{n-1}^2+x_n^2)}{n^2}=$$
$$=\frac{n(x_1^2+\ldots+x_n^2)}{n^2}=\frac{x_1^2+\ldots+x_n^2}{n},$$

что и требовалось.

Сопоставляя неравенство между средним геометрическим и средним арифметическим и доказанное неравенство (1), получаем, что *для любых неотрицательных чисел* $x_1,\ldots,x_n$ *имеет место точное неравенство*

$$\sqrt[n]{x_1\ldots x_n}\leqslant\left(\frac{x_1^2+\ldots+x_n^2}{n}\right)^{1/2}. \tag{3}$$

В частности, при $n=2$ получаем $\sqrt{x_1x_2}\leqslant\sqrt{\dfrac{x_1^2+x_2^2}{2}}$. Этому неравенству можно без труда придать геометрическое содержание. Из него, например, сразу следует, что *среди прямоугольников, вписанных в круг, наибольшую площадь имеет квадрат*. Эта задача, в свою очередь, допускает по меньшей мере два стереометрических обобщения. Одно из них такое: *среди прямоугольных параллелепипедов, вписанных в шар, найти параллелепипед наибольшего объема*. Другое обобщение: *среди цилиндров, вписанных в шар, найти цилиндр наибольшего объема*. Обе эти стереометрические задачи исследовал Кеплер. Об этом будет идти речь в следующем нашем рассказе. Отметим, кстати, что из неравенства (3) при $n=3$ немедленно следует, что *параллелепипедом наибольшего объема, вписанным в шар, является куб*. (Продумайте это!)

Планиметрическая задача о прямоугольнике наибольшей площади, вписанном в круг, также встретится нам дальше. Далее мы будем ее называть *планиметрической задачей Кеплера*.

5. Неравенство Коши — Буняковского. Имеет место следующая теорема:

для любых чисел $a_1,\ldots,a_n,b_1,\ldots,b_n$ *выполнено неравенство*

$$a_1b_1+\ldots+a_nb_n\leqslant(a_1^2+\ldots+a_n^2)^{1/2}(b_1^2+\ldots+b_n^2)^{1/2}. \tag{1}$$

Неравенство (1) называется *неравенством Коши—Буняковского*. Оно точное: равенство достигается при $a_1=b_1,\ \ldots,\ a_n=b_n$.

Докажем (1). Если $b_1 = \ldots = b_n = 0$, то доказывать нечего. Пусть не все b_i равны нулю. Имеем для любого числа x

$$(a_1 + xb_1)^2 + \ldots + (a_n + xb_n)^2 = a_1^2 + \ldots + a_n^2 + \\ + 2x(a_1b_1 + \ldots + a_nb_n) + x^2(b_1^2 + \ldots + b_n^2) = ax^2 + 2bx + c,$$

где введены обозначения

$$a = b_1^2 + \ldots + b_n^2, \quad b = a_1b_1 + \ldots + a_nb_n, \quad c = a_1^2 + \ldots + a_n^2.$$

Ясно, что $a > 0$ и что для всех x выполнено неравенство

$$ax^2 + 2bx + x \geqslant 0. \tag{2}$$

Условие неотрицательности квадратного трехчлена (2) состоит в неравенстве

$$b^2 - ac \leqslant 0 \iff (a_1b_1 + \ldots + a_nb_n)^2 \leqslant (a_1^2 + \ldots + a_n^2)(b_1^2 + \ldots + b_n^2), \tag{3}$$

что и требовалось.

Неравенство Коши—Буняковского имеет следующее важное обобщение.

6. Неравенство Гёльдера. Имеет место следующая теорема:

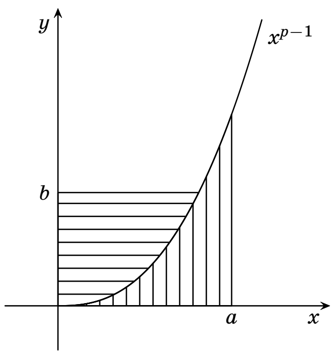
для любых неотрицательных чисел $a_1, \ldots, a_n$, $b_1, \ldots, b_n$ при $p > 1$ выполнено неравенство

$$a_1b_1 + \ldots + a_nb_n \leqslant (a_1^p + \ldots + a_n^p)^{1/p}(b_1^{p'} + \ldots + b_n^{p'})^{1/p'}, \tag{1}$$

где $p' = \dfrac{p}{p-1}$ (т. е. $\dfrac{1}{p} + \dfrac{1}{p'} = 1$).

Неравенство (1) называется *неравенством Гельдера*. Докажем его. Две функции $y = x^{p-1}$ и $x = y^{p'-1}$ взаимно обратны (проверьте). Выберем два положительных числа a и b. Тогда

$$\int_0^a x^{p-1}\,dx = \frac{a^p}{p}, \quad \int_0^b y^{p'-1}\,dy = \frac{b^{p'}}{p'}.$$



Теперь посмотрим на рис. 23. Величина a^p/p — это площадь криволинейного треугольника, заштрихованного вертикально, $b^{p'}/p'$ — площадь криволинейного треугольника, заштрихованного горизонтально. Легко проверить, что при любом расположении a и b сумма площадей этих двух треугольников не меньше площади прямоугольника со сторонами a и b. Причем равенство возможно, лишь если $a^{p-1}=b$.

Таким образом, доказано следующее неравенство для двух любых неотрицательных чисел:

$$ab \leqslant \frac{a^p}{p}+\frac{b^{p'}}{p'}. \tag{2}$$

Пусть теперь $a_1,\ldots,a_n$ и $b_1,\ldots,b_n$ — произвольные неотрицательные числа. Если, скажем, $b_1=\ldots=b_n=0$, то неравенство (1) верно. Значит, можно считать, что $A=(a_1^p+\ldots+a_n^p)^{1/p}\neq 0$ и $B=(b_1^{p'}+\ldots+b_n^{p'})^{1/p'}\neq 0$.

Положим $x_k=a_k/A$, $y_k=b_k/B$. Тогда из (2) следует

$$x_k y_k \leqslant \frac{a_k^p}{pA^p}+\frac{b_k^{p'}}{p'B^{p'}}, \quad k=1,2,\ldots,n.$$

Складывая эти неравенства и учитывая, что $1/p+1/p'=1$, $a_1^p+\ldots+a_n^p=A^p$, $b_1^{p'}+\ldots+b_n^{p'}=B^{p'}$, получаем

$$x_1y_1+\ldots+x_ny_n\leqslant 1 \implies a_1b_1+\ldots+a_nb_n\leqslant AB \implies$$
$$\implies a_1b_1+\ldots+a_nb_n\leqslant (a_1^p+\ldots+a_n^p)^{1/p}(b_1^{p'}+\ldots+b_n^{p'})^{1/p'},$$

что и требовалось.

Тема «Точные неравенства» необычайно обширна. Ей посвящено много книг и статей. Особенно известна книга Харди, Литлвуда и Полиа [19]. Эта тема пользуется большим успехом и в популярной литературе. Как правило, применение общих методов, о которых будет рассказано во второй части, дает возможность доказывать многие точные неравенства без труда. Но бывают и исключения. Вот две задачи, которые формулируются просто, но доказательство соответствующих экстремальных свойств в них, как мне кажется, не является простым. Попробуйте вы решить их. Возможно, вы найдете какие-то совсем нетрудные решения.

З а д а ч а 8. *Найти наименьшее значение суммы нечетного числа четвертых степеней величин* $x_1,\ldots,x_{2l+1}$, *если известно, что их сумма и сумма их кубов равна нулю, а сумма квадратов — единице.*

З а д а ч а 9. *Сто положительных чисел* $x_1,\ldots,x_{100}$ *удовлетворяют условиям* $x_1^2+\ldots+x_{100}^2>10000$, $x_1+\ldots+x_{100}<300$. *Доказать, что среди них найдутся три числа, сумма которых больше* 100.

Рассказ шестой

ЗАДАЧА КЕПЛЕРА

По обе стороны от места наибольшего значения убывание вначале нечувствительно.

И. Кеплер

«Когда историю жизни Кеплера сопоставляешь с тем, кем он стал и что он сделал, радостно изумляешься и при этом убеждаешься, что истинный гений преодолевает любые препятствия», — писал Гёте. Мне предстоит сейчас обрисовать одного из самых лучезарных, благородных и возвышенных гениев, которые когда-либо существовали. Не располагая возможностью сколько-нибудь подробно осветить эту необыкновенную личность, я отсылаю читателя к двум книгам о Кеплере. В книге Предтеченского «Кеплер, его жизнь и научная деятельность» [16] прекрасно отражена моральная высота Кеплера. В книге Белого «Иоганн Кеплер» [3] подробно рассказано о научном подвиге и гении Кеплера. Там же приведена обширнейшая библиография. Кроме этого, на русский язык переведены две книги самого Кеплера: «Стереометрия винных бочек» [11] и «О шестиугольных снежинках».

Какие только испытания не выпадали на долю Кеплера! Бедность, лишения, болезни и смерти близких, неустроенность, скитальчество. Но когда мы читаем Кеплера, то неизменно ощущаем его благодарность судьбе — за радость и счастье, дарованные ему. Радость от труда и счастье от постижения истины. Послушаем, как он сам говорит об этом. «Я предаюсь своему энтузиазму и не стесняюсь похвалиться перед смертными своим признанием: я похитил золотые сосуды египтян, чтобы создать из них храм моему богу вдали от пределов Египта. Если вы простите мне это — я порадуюсь, если укорите меня — снесу укор. Но жребий брошен. Я пишу эту книгу. Прочтется ли она моими современниками или потомством — мне нет до этого дела — она подождет своего читателя. Разве Господь Бог не ждал шесть тысяч лет созерцателя своего творения?» Так говорит он о своем третьем законе движения планет. А какую радость доставляют ему открытия других! «Я сидел дома, ничего не делал и думал о вас, глубокоуважаемый и славный Галилей, как вдруг узнал об открытии вами четырех планет при помощи телескопа... Я не мог без крайнего волнения подумать, что таким образом решился наш старинный спор... Может быть, я покажусь слишком смелым, если так легко поверю твоим утверждениям, не подкрепленным никаким собственным опытом. Но почему мне не верить ученейшему математику, о правоте которого свидетельствует сам стиль его суждений?» Так пишет он Галилею, восхищаясь его открытием спутников Юпитера. Снеллиуса он называет современным Аполлонием. Сталкиваясь с трудностями при решении геометрических задач, он обращается к нему так: «Представь же нам, Снеллий, гордость геометров нашего века, строгое решение этой и прочих задач, которые тут желательны». Он искренне верит в то, что любой человек, ищущий истину, будет счастлив, узнав о ее раскрытии. К таким людям адресованы его слова: «На некоторых местах... надо остановиться поподробнее, чтобы... ученые люди нашли, чем воспользоваться и чему порадоваться».

Предтеченский пишет: «Он всегда остается искренним и верным себе; его высокий ум был чужд честолюбия и тщеславия, он не искал от людей ни почестей, ни похвал... Он нисколько не выражает притязания на превосходство свое перед другими, теперь почти неизвестными учеными и во всю жизнь относится с глубоким почтением к Местлину, замечательному в наших глазах лишь тем, что ему посчастливилось иметь такого ученика, как Кеплер... Тихо Браге... был принципиальным противником его, так как не принимал горячо отстаиваемой Кеплером теории Коперника; мы знаем, что между обоими этими великими людьми были личные неприятности, и тем не менее Кеплер всегда восхваляет Тихо, отдавая ему должное и не пытаясь уменьшить его заслуг... Кеплер здесь, как и всегда, является искренним другом истины, что, к сожалению, так редко встречается... в наши дни».

С полным основанием Кеплер мог сказать о себе: «Я привык везде и всегда говорить правду».

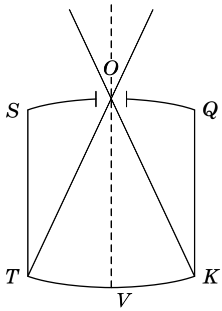
В русской литературе тема гения не нашла своего развития. Ее не касаются ни Толстой, ни Достоевский. Только Пушкин... Само слово «гений» произнесено в «Моцарте и Сальери», но чаще Пушкин пользуется еще более широким понятием — «Поэт».

Простодушие, верность дружеству, жажда творчества, способность радоваться всему прекрасному и, конечно, несовместность со злодейством — вот черты гения, которыми Пушкин наделил своего Моцарта и которыми в высшей степени обладал и Кеплер. И как Поэт, он неизменно следовал девизу: «Иди, куда влечет тебя свободный ум». Подобно тому как «Поэт сам избирает предметы своих песен», Кеплер избирал предметы своих научных изысканий.

Об одном таком «предмете» и будет наш рассказ.

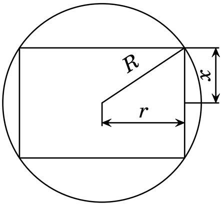
Вот как описывает Кеплер событие из своей жизни, случившееся осенью 1613 года, в книге «Стереометрия винных бочек».

«В ноябре прошлого года... я ввел в свой дом новую супругу в то время, когда Австрия, закончив обильный сбор благородного винограда, распределяла свои богатства... Весь берег в Линце был завален винными бочками, продающимися по сходной цене... Поэтому ко мне на дом было принесено и поставлено несколько бочек, а через четыре дня пришел продавец и промерил подряд все кадки, без различия, не обращая внимания на форму, без всяких соображений и вычислений. Именно, медный наконечник линейки просовывался через наливное отверстие полной бочки поперек до пятки того и другого деревянного круга, которые мы по-домашнему называем днищами, и после того, как в обоих случаях эта длина от верхней точки до нижней того и другого дощатого круга оказывалась равной, продавец объявлял количество амфор, вмещаемых бочкой, заметив лишь число на линейке в том месте, на котором оканчивалась заданная длина. Я удивился...» Кеплеру показалось странным, как с помощью одного измерения (для того чтобы его лучше понять, следует посмотреть рис. 24 из книги Кеплера) можно вычислить вместимость бочек разной формы. «Я, как новобрачный, счел для себя подходящим, — пишет далее Кеплер, — взять новый предмет математических занятий, и исследовать геометрические законы такого удобного в домашнем хозяйстве измерения, и выяснить его основания, если таковые имеются». Для выяснения «такого рода оснований» Кеплеру пришлось заложить основы дифференциального и интегрального исчисления, а заодно выдвинуть новые идеи для решения задач на максимум и минимум.



Ключевое место в книге «Стереометрия винных бочек» занимает теорема V части второй: «Из всех цилиндров, имеющих одну и ту же диагональ, самым большим и вместительным будет тот, в котором отношение диаметра основания к высоте равно $\sqrt{2}$».

Иначе говоря, в этой теореме дается решение следующей задачи: *вписать в заданный шар цилиндр наибольшего объема*. К нему естественно примыкает планиметрический вариант: *вписать в заданный круг прямоугольник наибольшей площади*.



Первую из сформулированных задач мы называем далее *задачей Кеплера*, вторую — *планиметрической задачей Кеплера*.

Сначала решим задачу Кеплера методом, которым решил бы ее (если бы только поставил) Тарталья. Пусть шар имеет радиус R. Половину высоты цилиндра обозначим через x (рис. 25). Тогда радиус r основания цилиндра равен $\sqrt{R^2 - x^2}$ и объем цилиндра равен $2\pi(R^2 - x^2)x$. А у Тартальи было, как мы помним, $(S^2 - x^2)x/4$. Тогда из формулы предыдущего рассказа получим, что максимальное $\widehat{x} = R/\sqrt{3}$, а $\widehat{r} = \sqrt{R^2 - \dfrac{R^2}{3}} = R\sqrt{\dfrac{2}{3}}$. Таким образом, отношение диаметра основания экстремального цилиндра к высоте равно $\sqrt{2}$.

Но Кеплер именно так и сформулировал, как мы видели, свой результат.

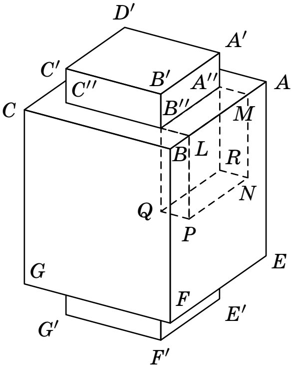
Кеплер мог бы воспользоваться своей же идеей о нечувствительности изменения функции вблизи максимума (см. эпиграф к данному рассказу). Но он прошел мимо этой возможности и дал чисто геометрическое решение.

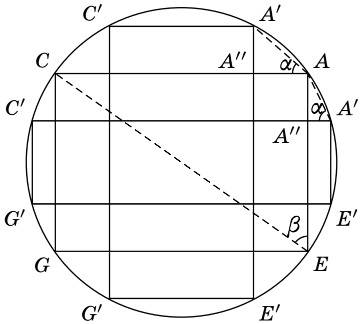
Вопрос о наиболее вместительном цилиндре, вписанном в шар, Кеплер сводит к решению следующей задачи на максимум: *из всех прямоугольных параллелепипедов с квадратными основаниями, вписанных в шар, куб имеет наибольший объем* *). Этот результат доказан в теореме IV части второй книги Кеплера.

Прямоугольные параллелепипеды с квадратными основаниями Кеплер называет коротко *столбами*, и мы последуем его примеру. Возможны два случая: а) столб выше куба; б) столб ниже куба.

Разберем сначала случай а) — см. рис. 26 и 27. Пусть куб $ABCDEFGH$ и «столб» $A'B'C'D'E'F'G'H'$ вписаны в одну и ту же сферу (на рис. 26 точки D, H, H' не видны). Сравним их объемы. Из куба «выпирают» два параллелепипеда с квадратными основаниями: сверху $A'B'C'D'A''B''C''D''$ и равный ему по объему нижний параллелепипед. При этом от куба «отнимается» гораздо больше. Усмотреть это совсем просто: у каждой стороны квадрата $A''B''C''D''$ прилегает к столбу параллелепипед с квадратным основанием, равным (конгруэнтным) самому квадрату $A''B''C''D''$. Один из таких параллелепипедов мы обозначили $A''B''QRMNPL$.

*) Это частный случай задачи, обсуждавшейся в предыдущем рассказе.





Уже эти четыре параллелепипеда занимают больший объем, чем выпирающие части столба. Действительно, объем выпирающих частей равен $2|A''B''|^2|A''A'|$, а объем четырех прилегающих параллелепипедов равен $4|A''B''|^2|A''M|$. Но $|A''M| = |A''A|/\sqrt{2}$. Теперь рассмотрим треугольник $A'AA''$ (рис. 27). Угол $\alpha = \widehat{A'AA''}$ опирается на дугу $A'C$, угол $\beta = \widehat{AEC}$ опирается на бóльшую дугу AC. Значит, $\alpha < \beta$, и, следовательно,

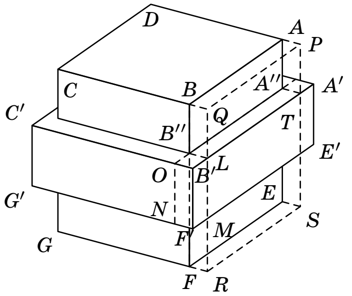
$$|A''A'| : |A''A| = \operatorname{tg}\alpha < \operatorname{tg}\beta = |CA| : |AE| = \sqrt{2}.$$

Получаем, что

$$2|A''B''|^2|A''A'| < 2|A''B''|^2|A''A|\sqrt{2} = 2|A''B''|^2\sqrt{2}\sqrt{2}|A''M| = \\ = 4|A''B''|^2|A''M|.$$

Нужное неравенство для объемов доказано.

Остается разобрать случай б) (рис. 28). Пусть снова куб $ABCDEFGH$ и столб $A'B'C'D'E'F'G'H'$ вписаны в одну и ту же сферу (точки H, D', H' на рис. 28 не видны). Сравним их объемы. Из столба выпирают два параллелепипеда с квадратными основаниями общим объемом $2|AB|^2|AA''|$. Объем части столба, выпирающей из куба, меньше. Кеплер доказывает это так. Приложим, говорит он, к каждой боковой стороне куба параллелепипед толщины, равной толщине выпирания столба. (На рис. 28 мы изобразили один из четырех таких параллелепипедов $ABFEPQRS$.) Их общий объем равен $4|AB|^2|AP| = 4|AB|^2|A''A'|/\sqrt{2}$. И снова, рассмотрев треугольники $AA'A''$ и AEC (на рис. 27),



устанавливаем, что угол α больше угла β, и значит, $|AA''| : |A''A'| > \sqrt{2}$. Но, как справедливо отмечает Кеплер, если прилепить к кубу четыре параллелепипеда, то они все-таки не закроют всего столба, будут «зиять» четыре параллелепипеда у ребер (один из них — $B''LB'OF''MF'N$ (точка F'' не видна) — мы изобразили на рис. 28). Каждый из этих параллелепипедов составляет часть столбиков, вытянувшихся у каждого из ребер AE, BF, CG и DH. Объем каждого такого столбика равен $|AB||AP|^2$. Но когда мы приложили к каждой грани параллелепипед, то они «выдаются за высоту столба восемью столбиками» (один из этих восьми — $BQPATA''B'L$ — нарисован на рис. 28). Объем каждого из восьми столбиков равен $|AB||AP||AA''|$. Очевидно, неравенство $2|AA''| > 2\sqrt{2}|A''A'| = 4|AP|$, из которого следует, что четыре столбика у ребер также уступают в объеме выделяющимся восьми параллелепипедам. Итак, «выпирает» из столба объем $2|AB|^2|AA''| > 4|AB|^2|AP|$, а из куба меньше, чем $4|AB|^2|AP| - 8|AB||AP||AA''| + 4|AB||AP|^2 < 4|AB|^2|AP|$. Итак, куб больше теряет, чем приобретает. Вспомогательная задача решена. (Вспомним, что даже несколько более общую задачу алгебраическими средствами мы исследовали в предыдущем рассказе.)

Остальное совсем просто. В каждый цилиндр можно вписать столб и отношение объема цилиндра к объему вписанного столба постоянно и равно (проверьте) $\pi/2$. Значит, наибольшим по объему является цилиндр, в который можно вписать куб. А у него отношение высоты к диаметру основания равно $\sqrt{2}$.

Доказав эту теорему, Кеплер пишет: «Отсюда ясно, что австрийские бочары как бы по здравому и геометрическому смыслу при построении бочки соблюдают правило, чтобы за радиус днища брать треть длины клепок. Именно при таком устройстве цилиндр, мысленно построенный между двумя днищами, будет иметь две половины, весьма близко подходящие к условиям теоремы V, и потому будет самым вместительным, хотя бы при постройке бочки от точных правил несколько и отступили, потому что фигуры, близкие к оптимальной, очень мало меняют свою вместимость..., ибо по обе стороны от места наибольшего значения убывание в начале несущественно».

В заключительных словах Кеплера заложен тот основной алгоритм нахождения экстремумов, который впоследствии был оформлен в точную теорему. Сначала (для многочленов) его описал Ферма (1629 г.), а затем — Ньютон и Лейбниц — в общем виде. Он впоследствии получил название «теоремы Ферма». Много интересного о задаче Кеплера читатель может почерпнуть из статьи М. Б. Балка «Секрет старого бондаря» (Квант, 1986, № 8, с. 14).

Рассказ седьмой

БРАХИСТОХРОНА

Если начальная и конечная точки движения одинаковы, то поскольку прямая есть кратчайшее расстояние между ними, то можно было бы думать, что движение, совершающееся по ней, требует наименьшего времени. На самом деле это не так.

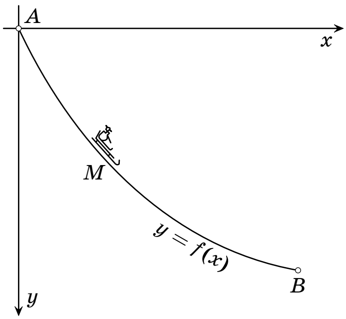
Г. Галилей

Невозможно отрицать глубокое значение, какое имеют точно поставленные проблемы для продвижения математической науки.

Д. Гильберт

С 1682 года стал выходить в свет научный журнал — «Акта Эрудиторум». В июньском номере этого журнала за 1696 год была помещена заметка знаменитого швейцарского ученого Иоганна Бернулли с интригующим заглавием: «Новая задача, к решению которой приглашаются математики».

Часто случается, что постановка новой проблемы привлекает внимание многих выдающихся ученых. Соревнуясь друг с другом, они создают мощные методы решения задач, которые потом щедро служат науке. Так вышло и с задачей И. Бернулли. Вот как формулирует ее сам автор.



В вертикальной плоскости даны точки A и B (рис. 29). *Определить путь AMB, спускаясь по которому под действием собственной тяжести, тело M, начав двигаться из точки A, достигнет точки B в кратчайшее время.*

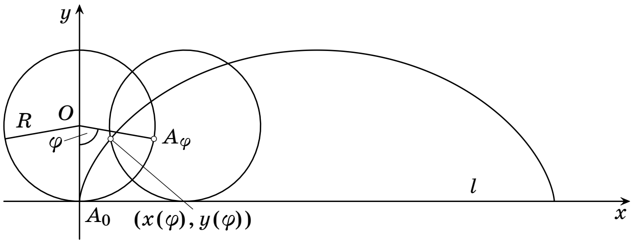
Ставя свою задачу, И. Бернулли не упоминает Галилея. И напрасно! Все современное естествознание «вышло» из Галилея. И дело не только в том, что он открыл основополагающие законы механики. Величайшая заслуга Галилея перед человечеством в том, что он первым стал *спрашивать* у Природы. Нынешний этап развития науки начался в тот момент, когда Галилей поднялся на Пизанскую башню для того, чтобы спросить у Природы о ее законах падения тел.

Галилей проводил эксперименты с наклонными плоскостями. По-видимому, он ставил также опыты и на круговых желобах. Вот что пишет он в «Беседах о механике» — главном труде всей своей жизни: «Тела, опускающиеся по дугам, соответствующим хордам, наклоненным к горизонту..., совершают движение, как показывает опыт, также в равные промежутки времени и притом меньшие, нежели движение по хордам». Из двух утверждений Галилея относительно движения по дугам верно лишь одно: движение по дуге быстрее, чем по хорде. Второе же утверждение — относительно равенства промежутков времени — верно лишь приближенно, и этот факт впоследствии теснейшим образом оказался связанным с задачей И. Бернулли.

Но так или иначе и приведенные выше слова Галилея, и его слова, взятые нами в качестве эпиграфа, сразу наталкивают на вопрос И. Бернулли: какая же кривая соответствует *кратчайшему* времени движения, т. е. какая же кривая является *брахистохроной* (по-гречески — наибыстрейшей)? Многие авторы упрекают Галилея в ошибке, утверждая, что он считал брахистохроной дугу окружности. В «Беседах» Галилей несколько раз возвращается к теме о сравнении движения по окружности и хорде, но ни в одном месте его слова нельзя истолковать так, что движение по дуге окружности — самое короткое среди всех кривых, соединяющих две заданные точки. Впрочем, может быть, какие-то его слова ускользнули от нашего внимания.

Однако пришло уже время вернуться к предмету нашего рассказа. Многие математики откликнулись на «приглашение» И. Бернулли. Одним из первых решивших задачу о брахистохроне был Лейбниц. Проблема ему очень понравилась, и он назвал ее прекрасной. Затем сообщили о своих успехах Якоб Бернулли (брат Иоганна) и Лопиталь. Сам Иоганн Бернулли, разумеется, располагал своим решением. Все эти ученые внесли значительный вклад в зарождавшееся новое направление — математический анализ. Но, кроме названных, было опубликовано и еще одно, безымянное, решение, в котором знатоки ex unge leonem («как по когтям узнают льва» — эту латинскую поговорку процитировал И. Бернулли) сразу же узнали Ньютона. Ньютон потом признался, что потратил на решение этой задачи около 12 часов непрерывного обдумывания.

Все авторы пришли к одному и тому же выводу: *брахистохроной является дуга циклоиды*. Здесь уместно будет сказать похвальное слово этой замечательной кривой.



Циклоиду описывает точка окружности, катящаяся без скольжения по прямой. Напишем ее уравнение.

Пусть l — горизонтальная прямая, по которой катится окружность радиуса R, имеющая центром точку O. Допустим, что в нулевой момент времени точка окружности, за которой надлежит следить, расположена в точке касания окружности с прямой l. Обозначим эту точку через A_0. Пусть в этот начальный момент времени A_0 совпадает с началом координат, где ось x направлена по прямой l, а ось y — в перпендикулярном направлении (рис. 30). Посмотрим, где окажется точка A_0 после того, как окружность повернется на угол φ по часовой стрелке. Для этого отложим точку A_φ на первоначальной окружности так, чтобы угол $A_\varphi O A_0$ был равен φ. Когда окружность при своем движении повернется на угол φ, точка A_φ займет положение точки касания. Таким образом, абсцисса центра нового положения окружности будет равна $R\cdot\varphi$, ибо такова длина дуги от A_0 до A_φ. При этом наша точка A_0 займет такое положение, что угол $A_0 O A_\varphi$ будет равен φ. Отсюда следует, что координатами $(x(\varphi), y(\varphi))$ точки A_0 будут числа

$$x(\varphi) = R(\varphi - \sin\varphi), \quad y(\varphi) = R(1 - \cos\varphi).$$

Вот мы и получили уравнение циклоиды, которая при $\varphi = 0$ проходит через начало координат. В общем случае появляется еще один параметр

$$x(\varphi) = R(\varphi - \sin\varphi) + C_1, \quad y(\varphi) = R(1 - \cos\varphi). \tag{1}$$

Что же такого примечательного в этой кривой? Как она возникла?

Циклоида впервые появилась в работах Галилея для иллюстративных целей. Он же дал ей название ц и к л о и д а, т.е. «связанная с кругом». Вскоре эту кривую переоткрыли во Франции (Мерсенн, Роберваль, Декарт, Ферма, Паскаль) и назвали *рулеттой* или *трохоидой*. Первое чудо, случившееся с этой кривой, состоит в том, что она служила как бы полигоном, на котором проверялись новые виды оружия, впоследствии поступившего в арсенал математического анализа.

Античная математика оставила будущим поколениям совсем мало кривых. Главные кривые античности — окружность, а также эллипс, гипербола и парабола, появившиеся у Аполлония. Счастье, что первые законы механики не вывели за пределы этого запаса кривых: планеты движутся по эллипсам, а брошенные тела летят по параболам.

Так вот, долгое время все крупнейшие математики XVII века (кроме названных, еще Вивиани, Торричелли и некоторые другие) оттачивали свои новые методы исследования именно на циклоиде: проводили касательные, находили площади под ней, вычисляли длину ее дуг и т. д.

А затем случилось второе чудо. Циклоида стала первой «неантичной» кривой, которая оказалась связанной с законами природы. Выяснилось, что именно циклоида, а не окружность, как писал Галилей, обладает тем свойством, что тело, скользящее по ней без трения, совершает колебания с периодом, не зависящем от начального положения. Это свойство циклоиды — ее *таутохронность* (т. е. равновременность) — открыл Гюйгенс. Его работа произвела подлинную сенсацию. Сам Гюйгенс писал об этом: «Наиболее желательным плодом, как бы величайшей вершиной этого учения Галилея о падении тел, является открытое мною свойство циклоиды».

И вот вновь циклоида появилась по совершенно другому поводу...

Теперь уже пора переходить к решению задачи. Мы помним, что имелись пять решений: И. Бернулли, Лейбница, Я. Бернулли, Лопиталя и Ньютона. Все они были очень содержательны. Лейбниц применил прием, который далее развил Эйлер (суть его можно понять из письма Лейбница к И. Бернулли, приводимого нами далее). Ныне метод Лейбница—Эйлера является одним из основных методов решения задач на максимум и минимум — это так называемый *прямой метод в вариационном исчислении*. Я. Бернулли основывал свое решение на принципе Гюйгенса и сделал, таким образом, еще один шаг к созданию теории Гамильтона—Якоби (о ней было кратко упомянуто в третьем рассказе). Но наибольшую популярность получило решение самого автора. Не счесть книг, где оно приведено. Приведем его и мы.

Введем в плоскости систему координат (x, y) так, чтобы ось x была горизонтальна, а ось y — направлена *вниз*. При этом поместим точку A в начало координат (рис. 29). Пусть уравнение кривой (желоба), соединяющей точку A с точкой B, координаты которой равны (a, b), будет задано функцией $y = f(x)$. Сначала нам надлежит вычислить время, за которое тело M массы m (без трения) спустится из точки A в точку B по желобу $f(x)$. Из механики известен закон Галилея, согласно которому скорость тела в точке с координатами $(x, f(x))$ (при движении под действием тяжести без трения) не зависит от формы кривой f между точками A и $(x, f(x))$, а зависит лишь от ординаты $f(x)$. «Скорости падающих весомых тел находятся между собою в отношении корней квадратных из пройденных высот», — пишет И. Бернулли.

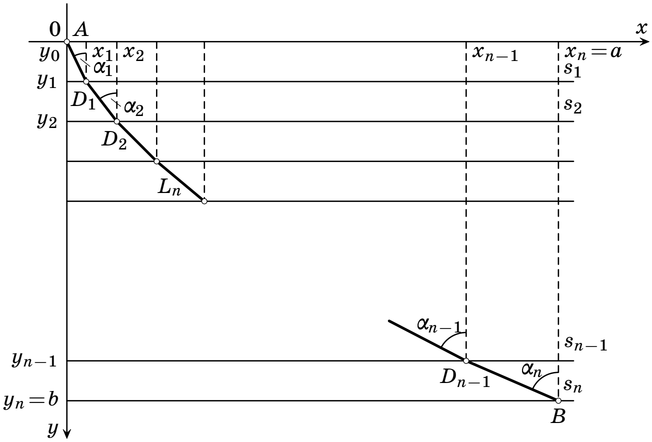
Действительно, кинетическая энергия нашего тела в точке $(x, f(x))$ равна $mv^2/2$ и эта энергия равна разности потенциальных энергий, т.е. $mgf(x)$. В итоге, скорость в точке $(x, f(x))$ оказывается равной $\sqrt{2gf(x)}$ (где g — ускорение силы тяжести). Рассмотрим теперь участок пути между точками $(x, f(x))$ и $(x+dx, f(x+dx))$, где dx — малое приращение абсциссы. Длина ds этого участка пути примерно равна $\sqrt{dx^2 + (f(x+dx) - f(x))^2}$. Воспользовавшись приближенным равенством $f(x+dx) - f(x) \approx \approx f'(x)\,dx$, получаем, что $ds = \sqrt{1 + (f'(x))^2}\,dx$. Скорость движения на малом участке можно считать постоянной и равной $\sqrt{2gf(x)}$. В итоге, время dt, требующееся для прохождения нашего малого участка, окажется примерно равным $dt \approx \frac{\sqrt{1+(f'(x))^2}}{\sqrt{2gf(x)}}dx$, а все время T движения от A до B запишется в виде интеграла

$$T = \int_0^a \frac{\sqrt{1+(f'(x))^2}}{\sqrt{2gf(x)}}dx. \qquad (2)$$

Таким образом, получается следующая аналитическая форма для задачи о брахистохроне. Требуется найти минимум интеграла $\int_0^a \frac{\sqrt{1+(f'(x))^2}}{\sqrt{2gf(x)}}\,dx$ по всем функциям f, для которых $f(0) = 0$, $f(a) = b$.

Мы записали нашу задачу на языке математики, в данном случае на языке интегрального исчисления. Эта процедура называется *формализацией* задачи. (Подробнее об этом будет сказано в десятом рассказе.) Элементы интегрального исчисления проходят теперь в школе, но в самом конце, а эта книга адресована более широкому кругу школьников. Поэтому объясним происходящее без ссылки на интегральное исчисление. Сделаем это в духе рассуждений математиков XVII века. Напомним при этом, что основные понятия анализа возникли всего за 12 лет до статьи И. Бернулли, и период «строгости» был еще далеко впереди.

Разделим отрезок $[0, b]$ оси ординат на n частей точками $0 = y_0$, y_1, y_2, ..., $y_n = b$ и найдем те абсциссы x_i, для которых $f(x_1) = y_1$, $f(x_2) = y_2$, ..., $f(x_{n-1}) = y_{n-1}$, $x_n = a$. Точки (x_i, y_i) и (x_{i+1}, y_{i+1}), где $i = 0, 1, \ldots, n-1$, соединим отрезками прямых. Тогда наряду с функцией $y = f(x)$ возникает ломаная L_n, соединяющая те же точки A и B (рис. 31). Эта ломаная «аппроксимирует» (т. е. приближает) функцию $y = f(x)$, и тем точнее, чем больше n. И время скольжения тела M по этой ломаной будет близко ко времени скольжения тела M по желобу, задаваемому функцией $y = f(x)$.



Скорость движения по $(i+1)$-му отрезку ломаной можно приближенно считать постоянной и равной $\sqrt{2gy_{i+1}}$, $i = 0, 1, \ldots, n$. При этом допущении время T_n движения по n-звенной ломаной можно вычислить точно:

$$T_n = \frac{\sqrt{y_1^2 + x_1^2}}{\sqrt{2gy_1}} + \frac{\sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}}{\sqrt{2gy_2}} + \ldots$$
$$\ldots + \frac{\sqrt{(y_n - y_{n-1})^2 + (x_n - x_{n-1})^2}}{\sqrt{2gy_n}}. \quad (2_n)$$

Это время требуется минимизировать.

Предел T_n при $n \to \infty$ и будет искомым временем движения тела M по кривой $y = f(x)$. Этот предел и есть тот интеграл (2), который был выписан выше. Но мы, следуя И. Бернулли, будем решать не задачу, сформулированную выше с помощью интегрального исчисления, а именно соответствующую (2_n) приближенную «дискретную» задачу. Приведем ее формулировку.

Даны две точки A и B, координаты которых $(0,0)$ и (a,b). Требуется на прямых $l_1,\ldots,l_n$, параллельных оси абсцисс и имеющих ординаты $y_1,\ldots,y_n$, найти точки $D_1 = (x_1y_1)$, ..., $D_{n-1} = (x_{n-1},y_{n-1})$ так, чтобы сумма T_n была минимальна (считая, что $x_0 = y_0 = 0$, $x_n = a$, $y_n = b$).

К этой задаче И. Бернулли применил замечательный прием, который оказал сильное влияние на всю последующую историю естествознания. О нем мы сейчас и расскажем.

Оптико-механическая аналогия. Вернемся немного назад, к третьему рассказу. Там мы поставили и решили задачу о преломлении света. Давайте внимательно вникнем в ее содержание и сравним его с задачей (2_n), где $n = 2$. В обоих случаях даны две точки и горизонтальная прямая и требуется найти точку на этой прямой так, чтобы минимизировать «взвешенную» сумму длин. Только в третьем рассказе скорости были произвольны — v_1 и v_2, и мы, по сути дела, минимизировали функцию

$$\frac{\sqrt{y_1^2 + x_1^2}}{v_1} + \frac{\sqrt{(a-x_1)^2 + (b-y_1)^2}}{v_2} \tag{3}$$

(если координаты точки $A = (0,0)$, точки $B = (a,b)$, а уравнение прямой l таково: $y = y_1$), а здесь имеется частный случай — мы должны минимизировать функцию

$$\frac{\sqrt{y_1^2 + x_1^2}}{\sqrt{2gy_1}} + \frac{\sqrt{(a-x_1)^2 + (b-y_1)^2}}{\sqrt{2gb}}. \tag{3'}$$

К этому же выводу пришел и Г. Лейбниц, а далее он пошел своим путем. Свой метод Лейбниц прекрасно выразил в письме к И. Бернулли (от 31 июня 1696 г.). Он пишет: «Мой метод несколько отличен от твоего, но, однако, приводит к тому же; для того чтобы, как требует справедливость, ответить на твою откровенность тем же, то вот он в немногих словах: заменив кривую многоугольником с бесконечно большим числом сторон, я вижу, что из всех возможных случаев (кривой) быстрейшего ската будет, если взять на ломаной три какие-нибудь точки, или вершины, A, C и B, причем точка C будет такой, что из всех точек, расположенных на горизонтальной кривой l, эта единственная дает скорейший путь от A к B. Таким образом, дело сводится к решению легкой задачи: даны две точки A и B и проходящая между ними горизонтальная прямая l; найти на этой прямой точку C, чтобы путь ACB был наискорейшим». (Мы несколько изменили обозначения Лейбница, чтобы подогнать их к своим: у него вместо C — B, вместо B — C, вместо l — прямая DE.)

И. Бернулли поступил по-другому. Представим себе неоднородную оптическую среду, состоящую из n однородных слоев $s_1, \ldots, s_n$ (посмотрим снова на рис. 31). Это можно представить себе, как n стекольных пластов, состоящих из разных стекол. Пусть при этом скорость распространения света в слое s_1 равна $\sqrt{2gy_1}$, в слое s_2 — $\sqrt{2gy_2}$, ..., а в слое s_n — $\sqrt{2gy_n}$. Каково же будет время распространения света, если его заставить идти по ломаной L_n? Очевидно, что T_n, где T_n задано формулой (2_n).

Таким образом, свет, согласно принципу Ферма, о котором говорилось в третьем рассказе, «решает» именно задачу (2_n) (если условиться, что скорость распространения света в i-м слое равна $\sqrt{2gy_i}$).

Отталкиваясь от механической задачи, мы вместе с И. Бернулли пришли к задаче оптической. Здесь впервые сработала *оптико-механическая аналогия*, которая затем принесла столько открытий в трудах Гамильтона, Якоби, Де Бройля и многих других.

Теперь уже можно перейти к решению И. Бернулли задачи о брахистохроне. Применим к задаче (2_n) в ее оптическом варианте закон Снеллиуса (о нем говорилось в третьем рассказе). Обозначим через α_i угол падения луча в i-м слое. В силу закона Снеллиуса получаем

$$\frac{\sin\alpha_1}{\sqrt{2gy_1}} = \frac{\sin\alpha_2}{\sqrt{2gy_2}} = \ldots = \frac{\sin\alpha_n}{\sqrt{2gy_n}} = \mathrm{const}. \tag{4}$$

Переходя к пределу при измельчении слоев, получим

$$\frac{\sin\alpha(x)}{\sqrt{2gf(x)}} = \mathrm{const}, \tag{5}$$

где $\alpha(x)$ — угол между касательной к кривой $y = f(x)$ в точке $(x, f(x))$ и осью Oy. Как известно, угловой коэффициент (т. е. тангенс угла, образованного касательной к графику функции f в точке x с осью Ox) равен $f'(x)$. Отсюда сразу следует, что

$f'(x) = \mathrm{tg}((\pi/2) - \alpha(x)) = \cos\alpha(x)/\sin\alpha(x)$, и значит,

$$\sin\alpha(x) = \frac{1}{\sqrt{1 + (f'(x))^2}}.$$

Из последнего равенства и (5) вытекает, что функция, график которой является решением задачи о брахистохроне, удовлетворяет соотношению

$$\sqrt{1 + (f'(x))^2}\sqrt{f(x)} = C,$$

где C — некоторая константа. Иначе говоря, функция должна удовлетворять дифференциальному уравнению

$$y' = \sqrt{\frac{C - y}{y}}. \tag{6}$$

А во времена И. Бернулли уже было известно, что уравнение (6) есть дифференциальное уравнение циклоиды.

Для тех, кто овладел некоторыми навыками интегрирования, произведем интегрирование уравнения (6).

$$y' = \sqrt{\frac{C - y}{y}} \iff \frac{\sqrt{y}\,dy}{\sqrt{C - y}} = dx.$$

Делаем подстановку $y = C\sin^2(t/2) = C(1 - \cos t)/2$. Тогда

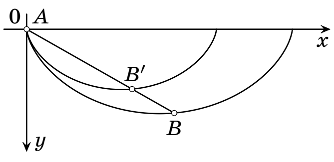
$$dx = \frac{\sqrt{C}\sin(t/2)\,d(C\sin^2(t/2))}{\sqrt{C}\cos(t/2)} = C\sin^2\frac{t}{2}\,dt = C(1 - \cos t)\frac{dt}{2}.$$

Интегрируя последнее соотношение, получим

$$x = \frac{C}{2}(t - \sin t) + C_1, \quad y = \frac{C}{2}(1 - \cos t). \tag{7}$$

Но это и есть уравнение циклоиды (1), следует лишь произвести замены букв: $C/2 \implies R$, $t \implies \varphi$.

Из полученных формул немедленно вытекает простой рецепт построения искомой циклоиды, т. е. той циклоиды, которая является решением задачи И. Бернулли. Отметим, что, все циклоиды (7) во-первых, гомотетичны, а во-вторых, выпуклы. Следовательно, можно взять любую из них, имеющую точку $A = (0, 0)$ своей левой вершиной (рис. 32), соединить точку $(0, 0)$ с точкой B прямой, которая пересечет построенную циклоиду в точке B', и затем следует сделать гомотетию с коэффициентом $|AB'|/|AB|$. Так получается искомая брахистохрона.



Решение задачи о брахистохроне доставило ее автору необыкновенную радость первооткрывательства. Вот как он говорит об этом: «Я не могу воздержаться от того, чтобы еще раз не выразить своего изумления по поводу отмеченного неожиданного тождества между гюйгенсовской таутохроной и нашей брахистохроной... Природа всегда действует простейшим образом, — так и в данном случае она с помощью одной и той же линии оказывает две различных услуги».

Метод Иоганна Бернулли дал возможность решить еще несколько замечательных задач из оптики, механики и геометрии. Приведем две такие задачи.

З а д а ч а 1. *Какова траектория световых лучей в атмосфере, где скорость распространения пропорциональна высоте?*

Этим вопросом задался Лопиталь, автор первого в истории учебника по математическому анализу. (Правда, недавно выяснилось, что в основной своей части этот учебник представляет собой обработку лекций, которые читал Лопиталю все тот же И. Бернулли.) Лопиталь сумел проинтегрировать уравнение и таким образом, ответить на поставленный вопрос. А по прошествии примерно двухсот лет выяснилось, что его решение имеет прямое отношение к геометрии Лобачевского: траектории световых лучей такие же, как *прямые Лобачевского в интерпретации Пуанкаре*.

З а д а ч а 2. *Найти минимальную поверхность вращения.*

Здесь надо сказать несколько слов, чтобы смысл задачи стал вполне ясен. Рассмотрим функцию $y = f(x)$, соединяющую две точки плоскости (x_0, y_0) и (x_1, y_1) и неотрицательную. Совершим ее обращение вокруг оси Ox. Получим *поверхность вращения*. Ее площадь задается формулой

$$S = 2\pi \int_{x_0}^{x_1} f(x)\sqrt{1 + (f'(x))^2}\, dx.$$

В задаче 2 требуется найти формулу кривой $y = f(x)$, при которой эта площадь минимальна. Задачу о минимальной поверхности вращения решили И. Бернулли и Лейбниц. Попробуйте и вы решить ее и задачу Лопиталя самостоятельно. А мы вместе обсудим их в четырнадцатом рассказе.

Задаче о брахистохроне суждено было сыграть выдающуюся роль в математическом анализе: она оказалась первой в ряду задач, из которых сформировалось вариационное исчисление.

Правда, незадолго до брахистохроны Ньютон рассматривал нечто похожее. Но об этом — в следующем рассказе.

Рассказ восьмой

АЭРОДИНАМИЧЕСКАЯ ЗАДАЧА НЬЮТОНА

Эта книга («Начала натуральной философии» Ньютона) навсегда останется памятником глубины гения.

П. Лаплас

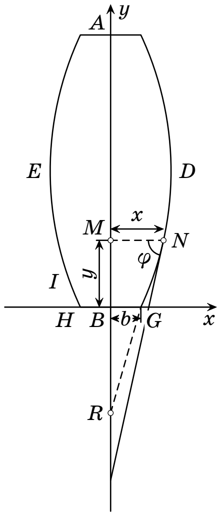
Ошибаются ли гении? Обычно, если задают такой вопрос, то предполагают утвердительный ответ на него. Есть нечто утешительное в сознании, что и гениям свойственно заблуждаться. Об ошибках гениев пишут с большим воодушевлением. Доставалось и Ньютону. В одной из книг, посвященных задачам оптимизации, написано так: «Ньютон сформулировал вариационную задачу о теле вращения, испытывающем наименьшее сопротивление в газе. Принятый им закон сопротивления физически абсурден, в результате чего поставленная им задача не имеет решения (чем более зазубрен профиль, тем меньше сопротивление)... Если бы выводы Ньютона были хотя бы приблизительно верны, то мы не нуждались бы сегодня в дорогостоящих экспериментах в аэродинамических трубах».

Что же это за задача и насколько справедливы приведенные выше критические слова? Об этом и пойдет речь в этом рассказе.

В 1687 году вышли «Математические начала натуральной философии» Ньютона. Никакое произведение научной литературы не может быть сопоставлено с этой книгой. В ней Ньютону было суждено открыть систему мира, а такое может случиться лишь однажды. Лагранж назвал это сочинение «величайшим из произведений человеческого ума»; «памятником глубины гения» назвал «Начала» Лаплас.

В этой книге, помимо открытия основных законов механики, законов движения планет и других основополагающих фактов, уделено место и многим частным проблемам.

Обсуждая вопросы, связанные с сопротивлением, оказываемым материальным телам средой, в которой они движутся, Ньютон как бы мимоходом бросил следующую фразу: «Когда же фигура $DNFG$ будет кривою такого рода, что если из любой ее точки N опустить на ось перпендикуляр NM и из заданной точки G провести прямую GR, параллельную касательной к кривой в точке N и пересекающую ось y в точке R, то имеет место пропорция $MN : GR = GR^3 : (4BR \cdot GB^2)$, тогда тело, образующееся при обращении этой кривой около оси AB, при движении в вышеупомянутой редкой среде в направлении от A к B будет испытывать меньшее сопротивление, нежели всякое иное тело вращения, описанное на той же длине и ширине» (рис. 33).



Фраза Ньютона привлекла к себе внимание современников лишь после того, как спустя девять лет, в 1696 году, И. Бернулли поставил свою задачу о брахистохроне, о которой речь шла в предыдущем рассказе.

В сиянии брахистохроны задача Ньютона заняла положение несчастной Золушки: ее как-то избегали, вспоминали о ней редко, да и то, как правило, чтобы поведать о заблуждении гения. На все это были свои причины. Но как для Золушки, так и для задачи Ньютона пришел свой час.

Попробуем проникнуть в замысел Ньютона.

При конструировании кораблей, снарядов, торпед или ракет естественно возникает стремление придать им такую форму, чтобы они испытывали возможно меньшее сопротивление при своем движении. Ньютон пишет: «Можно сравнить сопротивление тел между собой и находить те, которые наиболее приспособлены к продолжению свого движения в сопротивляющейся среде». Но, если, скажем, корабль или самолет не может иметь слишком большой симметрии своего корпуса, то головки ракет, снарядов и торпед естественно делать круглыми в поперечном сечении; иными словами, им резонно придавать форму *тел вращения*. Но какую именно? Сферическую, коническую, веретенообразную или еще какую-нибудь? На такие вопросы нельзя ответить без вычислений, без решения некоторой математической задачи на максимум и минимум. Именно такую задачу ставит (и решает в приведенной выше фразе) Ньютон.

В самом первом приближении задача ставится так.

Задача. *Найти тело вращения заданной длины и заданной ширины, испытывающее наименьшее сопротивление при движении в некоторой среде.*

Пояснения. Каждый из терминов, участвующий в формулировке (длина, ширина, движение и среда), требует точного описания.

Ньютон представляет себе тело вращения одинаковым сзади и спереди, т. е. симметричным относительно плоскости, проходящей через середину оси вращения и перпендикулярной этой оси. Таким образом, длина тела — это его длина по оси вращения, а радиус срединного сечения — это ширина тела. Из сказанного вытекает, что при всех рассмотрениях можно ограничиться лишь половинкой тела. Так мы дальше и поступаем.

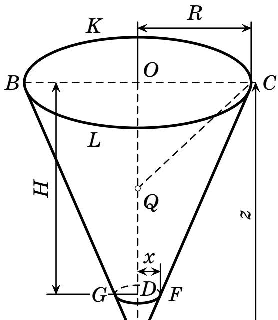
Теперь — о движении. Будем считать, вслед за Ньютоном, что движение, совершаемое телом, — это равномерное движение с постоянной скоростью v.

Наконец, о среде. Это наиболее тонкий и принципиальный вопрос. Среда, описанная Ньютоном, несколько необычна. Он сам называет ее «редкой» (разреженной). Ньютон представляет себе редкую среду «состоящей из равных частиц, свободно расположенных на равных друг от друга расстояниях». Неподвижные частицы имеют фиксированную массу m и являются абсолютно упругими шарами. Само тело Ньютон считает также абсолютно упругим, так что каждый шарик, столкнувшись с движущимся телом, отскакивает от него по закону «угол падения равен углу отражения».

После приведенных пояснений можно было бы сразу точно поставить и начать решать общую задачу, но мы поступим по-другому. Сначала исследуем более простой вопрос. С этого более простого вопроса, кстати говоря, начинает и сам Ньютон.

Задача об усеченном конусе. *При данных основании и высоте усеченного конуса найти тот конус, который испытывает наименьшее сопротивление при движении в редкой среде.*

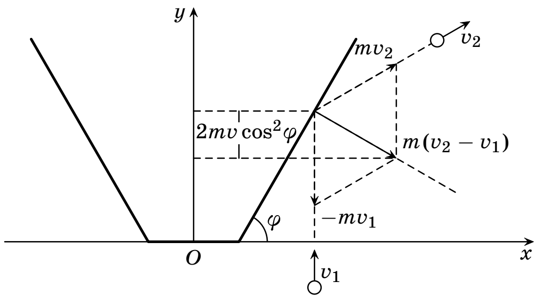
Определим силу сопротивления, испытываемую усеченным конусом при движении в редкой среде.



Пусть усеченный конус имеет высоту, равную H, и радиус верхнего основания R (рис. 34). Часть рис. 34, выполненная жирными линиями, содержится в «Математических началах».

Ньютон пишет: «Так как действие среды на тело то же самое... движется ли тело в покоящейся среде или же частицы среды ударяют с той же скоростью покоящееся тело, то будем рассматривать тело в покое». Поступим так же и мы: оставив конус неподвижным, мы будем считать, что на него снизу вверх «надвигается» среда со скоростью v.

Поверхность конуса, испытывающая столкновения с шариками среды, состоит из его нижнего основания и боковой поверхности. Сначала вычислим сопротивление, испытываемое нижним основанием. Обозначим его радиус через x. За единичное время с этим основанием столкнутся шарики, которые первоначально находились в цилиндре с основанием, равным нижнему основанию, и высотой, равной v. Объем этого цилиндра V_0 равен $\pi x^2 v$. Пусть ρ — плотность среды, m — масса одной частицы. Тогда число частиц, ударившихся о нижнее основание за единичное время, будет равно $N_0 = \frac{\rho}{m} V_0 = \frac{\rho}{m}\pi x^2 v$. Каждая частица после удара о нижнее основание сменит свою скорость на противоположную, т. е. получит приращение импульса, равное $-2mv$. По третьему закону Ньютона конус получит противоположное приращение импульса, и значит, общее приращение импульса от всех частиц будет равно $N_0 2mv = 2\pi\rho x^2 v^2$. Аналогично обстоит дело и для боковой поверхности. О нее ударяются частицы, помещаемые в объем между двумя одинаковыми боковыми поверхностями, и этот объем V_1 равен $\pi(R^2 - x^2)v$. Число частиц, ударившихся о боковую поверхность, равно $N_1 = \frac{\rho}{m} V_1 = \frac{\rho}{m}\pi(R^2 - x^2)v$. Но здесь надо аккуратнее подсчитать приращение импульсов. Отразившись от стенки, частица получит приращение импульса $m(v_2 - v_1)$ (рис. 35). Этот вектор надо спроектировать на ось y.



Из рисунка легко понять, что эта проекция равна $-2mv\cos^2\varphi$, где φ — угол наклона образующей конуса к плоскости нижнего основания. Таким образом, общее приращение импульса от всех частиц, ударившихся о боковую поверхность, равно

$$N_1 2mv\cos^2\varphi = 2\pi\rho(R^2 - x^2)v^2\cos^2\varphi.$$

Отметим (для будущего) формулу для силы сопротивления, испытываемого боковой поверхностью усеченного конуса, получающегося при вращении вокруг оси y отрезка $[AB]$, наклоненного к оси под углом φ, с абсциссами концов, равными a и b соответственно:

$$F = K(b^2 - a^2)\cos^2\varphi, \quad K = 2\pi\rho v^2. \tag{1}$$

В результате получаем следующее выражение для сопротивления, испытываемого всем конусом:

$$F(x) = K[x^2 + (R^2 - x^2)\cos^2\varphi], \quad K = 2\pi\rho v^2.$$

Выражение $\cos\varphi$ через x найти очень легко:

$$\cos\varphi = \frac{R - x}{\sqrt{(R-x)^2 + H^2}}.$$

Постоянный множитель K не влияет на поведение максимальных и минимальных величин, и мы его отбросим.

Теперь можно формализовать задачу об усеченном конусе.

З а д а ч а 1. *Найти минимум функции*

$$f(x) = x^2 + (R^2 - x^2)\frac{(R-x)^2}{(R-x)^2 + H^2} \tag{2}$$

при условии, что $0 \leqslant x \leqslant R$.

Перевод задачи на язык математики, как правило, не однозначен. И здесь, наряду с проведенной формализацией, можно предложить другое математическое описание той же задачи. Эта другая форма будет нам также очень полезна. Приведем ее.

Продолжим отрезок прямой, соединяющей точки C и F (рис. 34), до пересечения с осью y в точке S. Длину отрезка OS примем за новую переменную z. Из подобия треугольников SOC и SDF получаем: $x/R = (z - H)/z$, откуда $R - x = RH/z$ и $x = R(z - H)/z$. Подставив эти выражения в (2), получим новое выражение для силы сопротивления (деленной на K) через новую переменную:

$$g(z) = R^2\frac{(z - H)^2 + R^2}{z^2 + R^2}. \tag{3}$$

При этом z изменяется от H до бесконечности. Таким образом, мы пришли к следующей формулировке задачи об усеченном конусе (постоянный множитель R^2 снова отбрасывается).

З а д а ч а $1'$. *Найти минимум функции*

$$h(z) = \frac{(z-H)^2 + R^2}{z^2 + R^2}. \tag{$1'$}$$

при условии, что $z \geqslant H$.

Решим задачу $1'$. Эта задача получилась столь простой, что решить ее можно даже и без помощи дифференциального исчисления. При этом мы получим ответ в той же форме, как и у Ньютона.

Обозначим наименьшее значение функции h через m. Таким образом, $h(z) \geqslant m$ для любого $z \geqslant H$. При этом $m \leqslant h(H) = R^2/(R^2 - H^2) < 1$. Сделаем очевидные преобразования:

$$\begin{aligned}
&h(z) \geqslant m && \text{при } z \geqslant H \iff \\
\iff\ &(z-H)^2 + R^2 - mz^2 - mR^2 \geqslant 0 && \text{при } z \geqslant H \iff \\
\iff\ &z^2(1-m) - 2zH + H^2 + R^2(1-m) \geqslant 0 && \text{при } z \geqslant H
\end{aligned} \tag{4}$$

Если вдруг окажется, что существует такое $m < 1$, при котором неравенство (4) верно для всех z, и если при этом m можно найти $\widehat{z} \geqslant H$, которое обращает это неравенство в равенство, то это будет означать, что задача решена. Попробуем теперь отыскать нужные m и $\widehat{z}$. Если имеет место неравенство $az^2 + 2bz + c \geqslant 0$ для всех z и при этом $a\widehat{z}^2 + 2b\widehat{z} + c = 0$, то отсюда следует, что $D = b^2 - ac = 0$ и $\widehat{z} = -b/a$ (продумайте это). У нас $a = 1-m$, $b = -H$, $c = H^2 + R^2(1-m)$. Уравнение $D = 0$ приводит к соотношению

$$D = H^2 - (1-m)(H^2 + R^2(1-m)) \implies R^2m^2 - (2R^2 + H^2)m + R^2 = 0,$$

откуда

$$m = \frac{2R^2 + H^2 - H\sqrt{4R^2 + H^2}}{2R^2}.$$

Знак «+» был отброшен, ибо то m, которое мы ищем, должно быть, как это было установлено, меньше 1. Далее,

$$\widehat{z} = -\frac{b}{a} = \frac{H}{1-m} = \sqrt{R^2 + \left(\frac{H}{2}\right)^2} + \frac{H}{2} \quad (> H). \tag{5}$$

В итоге, нашлись такие $m < 1$ и $\widehat{z} > H$, что дискриминант уравнения $z^2(1-m) - 2zH + H^2 + R^2(1-m) = 0$ равен нулю и оно имеет единственный корень $\widehat{z} \geqslant H$, т. е. $h(z) \geqslant m$, $h(\widehat{z}) = m$. Задача $1'$ решена.

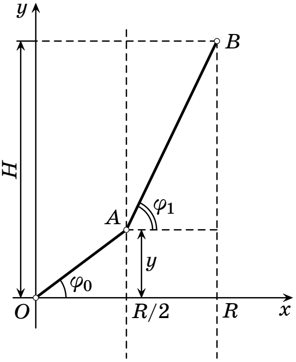
Приведем ответ задачи об усеченном конусе в той форме, которую придал ему Ньютон: «Если на круговом основании $CKBL$ (рис. 34), описанном из центра O радиусом OC, требуется построить такой усеченный конус $BCFG$ с высотою OD, коего сопротивление было бы меньше сопротивления всякого другого усеченного конуса, построенного на том же основании и высоте и движущегося по оси OD в сторону D, то, разделив высоту OD в точке Q пополам, продолжи OQ до S так, чтобы было $QS = QC$, S и будет вершиною искомого конуса, который усекается».

Легко убедиться, что геометрическая форма, приданная ответу Ньютоном, в точности соответствует аналитической форме, заключенной в формуле (5).

Получилось, что конус наименьшего сопротивления на самом деле усеченный, затупленный, а не заостренный. Ньютон при этом идет еще дальше. Он устремляет высоту H к нулю. При этом $\widehat{z}$ устремляется к R и угол при основании конуса устремляется к $45°$. Основываясь на этом, Ньютон делает замечание, что овальное тело следует заменять затупленным, т. е. делать спереди плоскую круглую площадку, образующую угол с примыкающей к площадке поверхности, равный $135°$. «*Я считаю, что это замечание,* — говорит Ньютон, — *может быть небесполезно при построении судов*».

З а д а ч а Н ь ю т о н а д л я д в у з в е н н о й л о м а н о й.

Сделаем еще один, предпоследний, шаг к решению общей задачи Ньютона. Будем рассматривать не любые кривые вращения, а лишь двузвенные ломаные с изломами на прямой $x = R/2$. То есть попробуем последовать методу Лейбница.



Точнее говоря, вопрос, который мы собираемся исследовать, состоит в том, чтобы на вертикальной прямой $x = R/2$ найти такую точку A, для которой поверхность, образованная вращением ломаной OAB (где B имеет координаты (R, H)) вокруг оси y (рис. 36), испытывала бы наименьшее возможное сопротивление в редкой среде Ньютона.

Пусть отрезок $[OA]$ наклонен к оси Ox под углом φ_0, а отрезок $[AB]$ — под углом φ_1. Обозначим ординату точки A через y. Воспользовавшись формулой (1), получаем следующее выражение для силы сопротивления, испытываемого телом вращения двузвенной ломаной OAB:

$$F = K\left[\left(\frac{R}{2}\right)^2 \cos^2\varphi_0 + 3\left(\frac{R}{2}\right)^2 \cos^2\varphi_1\right], \quad K = 2\pi\rho v^2.$$

При этом

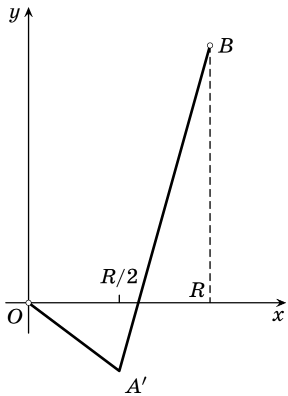
$$\cos\varphi_0 = \frac{R}{2\sqrt{y^2 + (R/2)^2}}, \quad \cos\varphi_1 = \frac{R}{2\sqrt{(H-y)^2 + (R/2)^2}}.$$

Таким образом, нам надлежит найти минимум функции

$$g_1(y) = \frac{1}{a^2 + y^2} + \frac{3}{a^2 + (H-y)^2},$$

где мы снова сократили на множитель K и обозначили $R/2$ через a.

Однако очень легко сообразить, что при неограниченном возрастании $|y|$ функция g_1 устремится к нулю. Вместе с тем она положительна. Значит, минимум этой функции равен нулю, хотя он и не достигается (подробнее аналогичный пример разбирается нами в одиннадцатом рассказе (п. 1)). Казалось бы, мы пришли к бессодержательной задаче... Но на самом деле из физических соображений вытекает, что нам надо найти минимум функции g_1 не по всем y, а лишь по y из отрезка $[0, H]$. Поясним это важное место.



Если взять $y < 0$, ломаная примет вид $OA'B$, изображенный на рис. 37. Тело вращения, образованное этой ломаной, будет иметь коническую воронку, и значит, некоторые шарики, образующие редкую среду, будут многократно отражаться от поверхности этой воронки. При этом закон сопротивления будет совсем иным, причем сопротивление возрастет. Таким образом, из физической сути задачи следует, что нельзя допустить, чтобы $y < 0$. Аналогично нельзя допустить, чтобы y был больше H. При постановке задачи неявно подразумевается *монотонность кривой вращения*. Так получается следующая формализация задачи о двузвенной ломаной.

З а д а ч а 2. *Найти минимум функции*

$$g_1(y) = \frac{1}{a^2 + y^2} + \frac{3}{a^2 + (H-y)^2}$$

при условии, что $0 \leqslant y \leqslant H$.

Решение задачи 2 нетрудно получить с помощью дифференциального исчисления.

Здесь мы ограничимся лишь формулировкой результата, который состоит в следующем:

а) существует такое $\delta > 0$, что при $0 \leqslant H \leqslant \delta$ минимум достигается в нуле;

б) для $H > \delta$ минимум достигается во внутренней точке отрезка $[0, H]$. При этом выполняется следующее равенство:

$$\frac{R}{4}\operatorname{tg}\varphi_0\cos^4\varphi_0 = 3\frac{R}{4}\operatorname{tg}\varphi_1\cos^4\varphi_1.$$

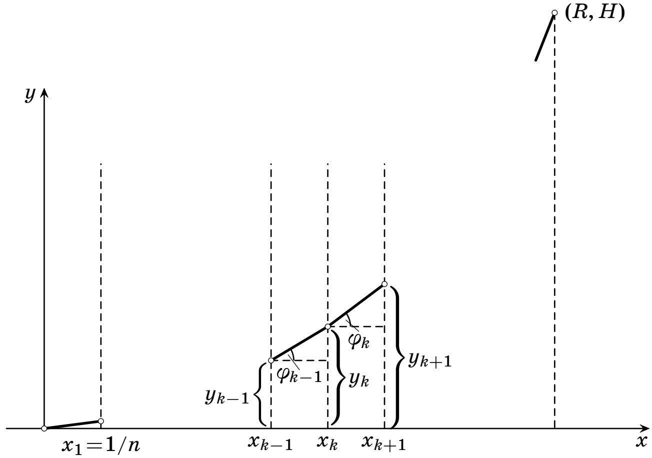
Таким образом, при достаточно малом H тело вращения, образованное двузвенной ломаной, будет затупленным (усеченным конусом), при больших H — заостренным (его поверхность будет состоять из конической поверхности и боковой поверхности усеченного конуса).

Сделаем теперь последний промежуточный шаг.

Задача Ньютона для n-звенной ломаной и ее формализации. Рассмотрим совокупность вертикальных прямых $l_1, l_2, \ldots$, имеющих уравнения

$$x = \frac{1}{n}, \quad x = \frac{2}{n}, \quad \ldots$$

(рис. 38), и ещё вертикальную прямую $x = R$ (если $R \neq k/n$ для некоторого k). Рассмотрим совокупность всех монотонно возрастающих ломаных с изломами на прямых l_k, соединяющих точку $(0, 0)$



с точкой (R, H). Попробуем среди всех описанных ломаных найти такую, чтобы поверхность, образованная вращением этой ломаной вокруг оси y, испытывала бы наименьшее сопротивление в редкой среде.

Пусть ордината вершины ломаной, расположенной на прямой l_k, равна y_k; угол наклона звена этой ломаной, соединяющей вершины, расположенные на l_k и l_{k+1}, к оси x обозначим через φ_k. Из формулы (1) выведем следующее выражение для силы сопротивления, испытываемого телом, полученным от вращения нашей ломаной, в редкой среде (для простоты считаем, что $R = nN$):

$$\begin{gathered}F = K\left\{\frac{1}{n^2}\cos^2\varphi_0 + \left[\left(\frac{2}{n}\right)^2 - \left(\frac{1}{n}\right)^2\right]\cos^2\varphi_1 + \ldots\right.\\ \left.\ldots + \left[\left(\frac{N}{n}\right)^2 - \left(\frac{N-1}{n}\right)^2\right]\cos^2\varphi_N\right\} =\\ = \frac{K}{n^2}[\cos^2\varphi_0 + 3\cos^2\varphi_1 + \ldots + (2N-1)\cos^2\varphi_N],\\ K = 2\pi\rho v^2, \quad \cos\varphi_k = \frac{1}{n}\frac{1}{\sqrt{1/n^2 + (y_{k+1} - y_k)^2}}.\end{gathered} \tag{7}$$

Таким образом, задача Ньютона для N-звенной ломаной состоит в том, чтобы минимизировать силу F по всем возможным наборам чисел $(y_1, \ldots, y_N)$, где $0 \leqslant y_1 \leqslant y_2 \leqslant \ldots \leqslant y_{N-1} \leqslant H$.

Для решения этой задачи используем опыт, полученный при исследовании задачи 2, ибо по сути дела задача для N-звенной ломаной как бы распадается на $N-1$ задачу типа задачи 2. Действительно, допустим, что нам удалось решить задачу об N-звенной ломаной и $(\widehat{y}_1, \ldots, \widehat{y}_{N-1})$ — ординаты вершин минимальной ломаной. Зафиксируем все вершины минимальной ломаной, кроме k-й вершины. При этом будем искать минимальное значение сопротивления, изменяя положение k-й вершины, иначе говоря, рассмотрим задачу типа задачи 2, когда требуется найти минимум функции

$$\begin{aligned}g_k(y) &= \frac{1}{n^2}[(2k-1)\cos^2\varphi_{k-1} + (2k+1)\cos^2\varphi_k] =\\ &= \frac{1}{n^4}\left[\frac{2k-1}{\dfrac{1}{n^2} + (y - \widehat{y}_{k-1})^2} + \frac{2k+1}{\dfrac{1}{n^2} + (\widehat{y}_{k+1} - y)^2}\right]\end{aligned}$$

при условии, что $\widehat{y}_{k-1} \leqslant y \leqslant \widehat{y}_{k+1}$. Функция g_k составлена из двух слагаемых — $(k-1)$-го и k-го, входящих в выражение для силы F (см. (7)), с учетом того, что ордината k-й вершины не закреплена и обозначена через y (множитель K мы, как обычно, опускаем). Решение поставленной задачи, очевидно, есть $\widehat{y}_k$.

Задача о минимизации функции g_k подобна задаче 2, и для нее справедливы те же заключения, которые мы сделали для задачи 2. А именно:

а) существует такое число $\delta_k > 0$, что при $\widehat{y}_{k+1} - \widehat{y}_{k-1} \leqslant \delta_k$ минимум достигается при $\widehat{y}_k = \widehat{y}_{k-1}$;

б) для $\widehat{y}_{k+1} - \widehat{y}_{k-1} > \delta_k$ имеет место равенство

$$\left(k - \frac{1}{2}\right) \operatorname{tg} \varphi_{k-1} \cos^4 \varphi_{k-1} = \left(k + \frac{1}{2}\right) \operatorname{tg} \varphi_k \cos^4 \varphi_k. \tag{8}$$

Из утверждений а), б) сразу вытекает, что экстремальная ломаная может иметь лишь следующий вид: сначала она идет какое-то время по оси x: $\widehat{y}_1 = \ldots = \widehat{y}_s = 0$, а затем поднимается вверх, следуя правилу (8). А именно, из соотношения

$$Y = \left(s + \frac{1}{2}\right) \operatorname{tg} \varphi_s \cos^4 \varphi_s = \left(s + \frac{3}{2}\right) \operatorname{tg} \varphi_{s+1} \cos^4 \varphi_{s+1}$$

определяется φ_{s+1}, из соотношения

$$Y = \left(s + \frac{3}{2}\right) \operatorname{tg} \varphi_{s+1} \cos^4 \varphi_{s+1} = \left(s + \frac{5}{2}\right) \operatorname{tg} \varphi_{s+2} \cos^4 \varphi_{s+2}$$

определяется φ_{s+2} и т. д.

Выпишем выражение силы сопротивления N-звенной ломаной непосредственно через ординаты вершин ломаной. Для этого в формуле (7) сделаем замену:

$$\cos \varphi_k = \frac{n^{-1}}{\sqrt{1/n^2 + (y_{k+1} - y_k)^2}}.$$

Получим после очевидных преобразований:

$$F = 2K \left(\sum_{k=1}^{N-1} \frac{k - 1/2}{n} \frac{1}{1 + \left(\dfrac{\Delta y_k}{\Delta x}\right)^2} \Delta x \right).$$

Здесь $\Delta y_k = y_{k+1} - y_k$, $\Delta x = 1/n$.

Вспомнив понятие интегральной суммы и заметив, что $\Delta y_k / \Delta x_k \approx f'(x_k)$, нетрудно поверить, что при неограниченном увеличении числа N написанная сумма стремится к интегралу:

$$F = 2K \int_0^R \frac{x\,dx}{1+(f'(x))^2}. \tag{8$'$}$$

Р е ш е н и е з а д а ч и Н ь ю т о н а. Можно доказать, что при увеличении N минимальная N-звенная ломаная будет стремиться к минимальной кривой — кривой, дающей решение задачи. Это означает (если вспомнить описание минимальной N-звенной ломаной), что минимальная кривая устроена следующим образом: *экстремальная функция сначала идет по оси x: $f(x) = 0$, $0 \leqslant x \leqslant a$; затем она поднимается вверх по некоторой кривой* (*кривой Ньютона*). При этом выполняется следующее соотношение:

$$x \operatorname{tg} \varphi(x) \cos^4 \varphi(x) = \text{const}, \tag{9}$$

где $\varphi(x)$ — угол, образуемый касательной к графику функции $y = f(x)$ в точке $(x, f(x))$. Именно в соотношение (9) перейдут равенства (8), полученные нами при решении задачи об N-звенной ломаной.

Преобразуем соотношение (9). Мы знаем (из геометрического смысла производной), что $\operatorname{tg}\varphi(x) = f'(x)$, откуда следует, что $\cos\varphi(x) = [1+(f'(x))^2]^{-1/2}$. Таким образом, получаем

$$\frac{xf'(x)}{[1+(f'(x))^2]^2} = \text{const}. \tag{10}$$

Соотношение (10) есть *дифференциальное уравнение для кривой Ньютона*.

И, наконец, последнее. В момент a, когда горизонтальная кривая переходит в кривую Ньютона, производная кривой Ньютона равна единице (вспомним замечание Ньютона: «Небесполезное при построении судов»). Дифференциальное уравнение (10) совместно с условием, что производная в момент излома кривой равна единице, позволяет найти уравнение самой кривой Ньютона. В четырнадцатом рассказе мы это интегрирование проведем, а пока просто напишем вид кривой Ньютона. Координаты (x, y) кривой Ньютона $y = f(x, c)$ выражаются по формулам:

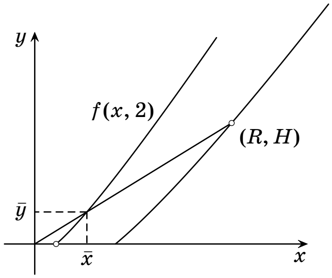
$$x = c\left(\frac{1}{u} + 2u + u^2\right), \quad y = c\left(\ln\frac{1}{u} + u^2 + \frac{3}{4}u^4\right) - \frac{7c}{4} \tag{11}$$

через параметр u. Константа c определяется из условия $f(c) = H$.

Пришла пора подводить итоги. Нам надлежит ответить на несколько вопросов.

1. Как же все-таки решить задачу Ньютона? Как же построить ту кривую заданной длины и ширины, вращение которой вокруг оси будет образовывать поверхность тела, испытывающего наименьшее сопротивление в редкой среде?

Из выражения (11) видно, что совокупность всех кривых, подозреваемых на минимум, зависит от одного параметра и при этом все кривые подобны друг другу. Значит, следует поступить так, как уже было с брахистохроной. Нарисуем кривую при некотором фиксированном c, скажем, при $c = 2$ (см. рис. 39). Можно показать, что кривая $f(x, 2)$ пересекается с любой прямой $y = kx$, $k > 0$, лишь один раз.



Для решения задачи Ньютона следует провести прямую $y = (H/R)x$, соединяющую начало координат с точкой (R, H). Пусть эта прямая пересечет график функции $y = f(x, 2)$ в точке $(\bar{x}, \bar{y})$. Тогда, положив $\bar{c} = H/\bar{y}$, построим кривую $y = f(x, \bar{c})$, которая пройдет через нужную точку (R, H).

Эта кривая в семействе (11) — единственная обладающая таким свойством. Она и будет решением задачи, поставленной Ньютоном.

2. Что же означает таинственная фраза Ньютона, которую мы процитировали в начале этого рассказа? Какое отношение имеет она к описанному выше решению задачи Ньютона?

На рис. 33 нанесены некоторые нужные нам сейчас обозначения. Мы обозначили $|MN| = x$, $|MB| = y$, $|BG| = b$, угол между отрезком $[MN]$ и касательной к кривой в точке N — через φ. Этот угол равен углу HGR. Далее, $\operatorname{tg}\varphi = f'(x)$. Значит,

$$|BR| : |BG| = \operatorname{tg}\varphi \implies |BR| = bf'(x),$$

откуда

$$|GR|^2 = |BG|^2 + |BR|^2 = b^2[1 + (f'(x))^2].$$

Теперь из пропорции Ньютона

$$|MN| : |GR| = |GR|^3 : (4|BR| \cdot |GB|^2)$$

получаем

$$\frac{x}{b\sqrt{1 + (f'(x))^2}} = \frac{b^3(1 + (f'(x))^2)^{3/2}}{4bf'(x)b^2} \iff \frac{xf'(x)}{(1 + (f'(x))^2)^2} = \frac{b}{4}. \qquad (12)$$

Получилось дифференциальное уравнение для кривой Ньютона (ср. (10) и (12)).

Затупленность тела вращения и излом в точке G (где угол, как мы помним, равен $135°$) были также предусмотрены Ньютоном. А отвечая на первый вопрос, мы установили, что дифференциальное уравнение (12) условие в изломе при данных ширине и длине определяют искомую кривую однозначно. Таким образом, можно считать, что *аэродинамическая задача была решена Ньютоном полностью*.

3. Почему же кривая Ньютона на протяжении чуть ли не трехсот лет была в положении несчастной Золушки? Почему так долго замысел Ньютона был не до конца понят?

Выше было сказано, что задача о брахистохроне открыла новую эру — эру вариационного исчисления. Вариационное исчисление бурно развивалось в течение почти двух столетий. Но, как уже говорилось, в наши дни многие технические задачи (в основном космического содержания) обнаружили, что часть актуальных проблем не может быть исследована методами вариационного исчисления. Потребовался новый шаг вперед, и была разработана новая теория — теория оптимального управления, — которая включила в себя вариационное исчисление и позволила решать задачи нового типа. Так вот, *задача Ньютона относится именно к оптимальному управлению*. В рамках этой теории задача Ньютона получает естественное и стандартное разрешение. А в рамках вариационного исчисления такого естественного и стандартного решения этой задачи не получалось. Таким образом, *Ньютон своей задачей просто опередил время почти на 300 лет*!

4. А что это за редкая среда? Бывает ли она? Не является ли абсурдным то обстоятельство, что тело наименьшего сопротивления затуплено? Кто это видывал торпеды или ракеты с затупленными головками?

Действительно, ни вода, ни окружающий нас воздух, ни любая другая привычная нам жидкая или газообразная среда не обладают свойствами редкой среды Ньютона. Так что решение Ньютона не может быть полезно при строительстве моторных лодок, катеров и океанских лайнеров. Однако физические допущения Ньютона и сама его аэродинамическая задача оказались на самом переднем крае современной науки в середине 50-х годов — в тот момент, когда настала эра сверхзвуковых и сверхвысотных летательных аппаратов. Там, «вверху», среда является «редкой». И *замечание Ньютона о затупленных конусах оказалось-таки «небесполезным» при построении судов, а именно космических и сверхзвуковых*.

И, наконец, зададим вопрос:

5. Ошибаются ли гении? В своей статье в журнале «Квант» (1982, № 5. С. 11—18) я оставил этот вопрос без ответа. Но когда в редакции статью прочитали Андрею Николаевичу Колмогорову, он потребовал, чтобы было написано: «Конечно, ошибаются». Да, может быть, ошибаются.

Но вот неожиданный пример — задача Ньютона. Двести пятьдесят лет казалось правдоподобным, что она не имеет физической базы, а ее решение абсурдно. Но «заблуждение» гения на поверку оказалось его прозрением.

Словом, не будем торопливы: *может статься, что мысль Гения, кажущаяся нам заблуждением, на самом деле несет в себе отпечаток истины — доступной ему, но еще не открывшейся нам*.

МЕТОДЫ РЕШЕНИЯ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ

Следует поставить перед собой цель изыскать способ решения всех задач... одним и притом простым способом.

Ж. Даламбер

Рассказ девятый

ЧТО ТАКОЕ ФУНКЦИЯ?

Это общее понятие требует, чтобы функцией от x называть число, которое дается для каждого x и вместе с x постепенно изменяется. Значение функции может быть дано аналитическим выражением, или условием, которое подает средства испытывать все числа и выбирать одно из них, или, наконец, зависимость может существовать и оставаться неизвестной.

Обширный вид теории допускает существование зависимости только в том смысле, чтобы числа, одно с другим в связи, принимать как бы данными вместе.

Н. И. Лобачевский

Прежде чем приступать к рассказам о методах исследования задач на максимум и минимум, поговорим о функциях.

Функция — основное понятие математического анализа. Но формировалось оно не сразу. Вначале оно было расплывчатым и не имело сколько-нибудь четкого описания. Первые попытки очертить контуры этого понятия делаются в конце XVII века Лейбницем, Иоганном и Якобом Бернулли. Сам термин «функция» принадлежит Лейбницу. И. Бернулли вкладывает в этот термин понятие о «*выражении, составленном каким-либо образом из переменной величины и постоянных величин*». Конкретизируя мысль И. Бернулли, Эйлер в своих учебниках определял функции как *аналитические выражения*, составленные из переменной величины и постоянных величин. Он же ввел символ $f(x)$.

Однако Эйлер считал возможным называть функцией также «произвольно начерченную кривую». Но Эйлеру были известны также и такие случаи, когда функция имела словесное описание. Скажем, тот же синус.

Рассмотрим один простой пример функции, допускающей разные описания. Это $y = |x|$. Несомненно, что $|x| = \sqrt{x^2} = (x^2)^{1/2}$ есть (аналитическое) выражение, составленное из переменной величины x и постоянных величин. Так что это — функция и по И. Бернулли, и по Эйлеру. Но ее же можно изобразить и как «начерченную кривую», идущую по биссектрисам первого и второго квадранта в декартовой системе координат. Ее же можно описать и словесно, сказав, что для x, равного нулю, эта функция равна нулю, для положительного x она равна самому x, для отрицательного x она равна $-x$. И вообще большинство функций допускают разные описания.

Чему же отдать предпочтение? По этому поводу было много споров. Скажем, Эйлер считал, что класс функций, являющихся «произвольно начерченными кривыми», шире, чем класс функций, задаваемых «аналитическими выражениями». Даламбер возражал ему, он утверждал, что получится одно и то же.

Д. Бернулли выдвинул, казалось бы, совсем парадоксальную точку зрения на функцию: он высказывал убеждение, что произвольную *периодическую* функцию с периодом 2π можно представить в виде суммы

$$\sum_{k=0}^{\infty} (a_k \cos kx + b_k \sin kx).$$

Большинству же казалось, что это — лишь некоторый узкий класс функций, что он у́же, чем «аналитические выражения» и, разумеется, чем «произвольные кривые».

В начале XIX века стала выкристаллизовываться идея функции как *соответствия*, закона, по которому независимая переменная x преобразуется в y, вне зависимости от способа такого соответствия. Среди первых ученых, пропагандировавших такую точку зрения, был Лобачевский. Его слова по этому поводу были вынесены нами в эпиграф к нашему рассказу.

Сходные воззрения и примерно в те же времена высказывались и во французской математической школе, и в немецкой. В начале века в учебниках по математическому анализу стали приводиться такие, например, определения функции.

«Пусть даны две переменные x и y, между значениями которых существует какая-нибудь зависимость. Вообще, одну из переменных, например x, рассматривают как независимую. Значение x может быть взято по произволу, но при данном x значение y не произвольно. Говорят тогда, что y есть функция от x» (Валле-Пуссен). Или: «Переменная величина y называется функцией от переменной величины x, если каждому значению x соответствует одно определенное значение y» (Немыцкий, Слудская, Черкасов).

Однако людей, особо приверженных к логической строгости (вообще говоря, таких людей не так уж много), подобные определения удовлетворить не могут. Ведь в них термин «функция» определяется через понятия, которые столь же неопределенны и столь же расплывчаты («зависимость», «закон», «соответствие» и т. п.).

Некоторое успокоение пришло с созданием теории множеств, начала которой были заложены в конце XIX века Георгом Кантором. Там вроде бы все стало на свои места. Пусть X и Y — два множества. Множество F пар (x, y), $x \in X$, $y \in Y$, называется функцией, если для любого $x \in X$ существует единственное y такое, что $(x, y) \in F$. Тогда пишут $y = f(x)$ *). Концепции теории множеств произвели огромное впечатление на многих математиков, которые были свидетелями зарождения новой теории. Гильберту, одному из величайших математиков всех времен, принадлежат следующие слова о теории множеств: «Я считаю, что она представляет собой высочайшее проявление математического гения и одно из самых высоких достижений чисто духовной деятельности человека». Нужно добавить при этом, что практически все современные математические работы так или иначе используют основные понятия теории множеств и ее символику.

Но увы, через какое-то время стали раздаваться критические голоса. Обнаружились противоречия, и снова возникли жаркие споры. В начале XX века все крупнейшие математики (Пуанкаре, Гильберт, Адамар, Вейль, Брауэр и другие) приняли участие в обсуждении проблем, связанных с кризисом основ всей математики. Для некоторых математиков теоретико-множественные определения (и в частности определение функции) были неприемлемо широки. Они высказывали убеждение в том, что

*) Что-то очень близкое по смыслу содержится в последней фразе цитаты из Лобачевского, приведенной нами в качестве эпиграфа.

функциональная зависимость, представляющая интерес с практической точки зрения, неизбежно должна быть «конструктивной». Это значит, что должно быть отчетливое правило (или, как теперь принято говорить, — алгоритм), которое по данному x позволяет отыскивать нужный y. В итоге возникли целые направления в математике, отрицавшие теорию множеств. Стала разрабатываться особая «конструктивная» математика, похожая и непохожая на уже ставшую привычной для всех математику, основанную на теории множеств.

Многие ученые очень тяжело переживали кризис, связанный с отрицанием частью математиков теоретико-множественных концепций. Давид Гильберт так писал об этом: «Никто не изгонит нас из рая, созданного для нас Кантором». Иные же горячо оспаривали эту точку зрения.

Однако пора вернуться к нашим проблемам: что же такое функция?

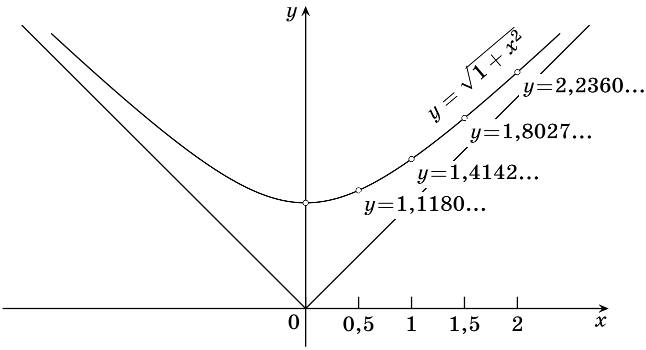
Какому воззрению отдать предпочтение? Где истина? Углубленное продумывание этих вопросов увело бы нас слишком далеко. Но необходимо сказать, что для «общеупотребительных» целей, для решения реальных задач, для физики и более широко — для приложений математики в естественных науках, в технике, экономике и т. п. годятся все достаточно определенные описания понятия функции.

Практически всюду в этой книге для наших основных целей достаточно будет того понимания, которое начало оформляться у самых истоков математического анализа. Поэтому мы будем трактовать понятие функции «по Бернулли» как «выражение, составленное каким-либо образом из переменной величины и постоянных величин».

Функции различаются по числу переменных. Сначала поговорим о функциях одной переменной, которая является действительным числом. Если мы будем говорить о функции $y = f(x)$, мы всегда будем иметь в виду какое-то точное предписание, следуя которому из числа x получится число y. По типу такого: «возьмите число, возведите его в квадрат, прибавьте к этому квадрату единицу и извлеките квадратный корень».

Так описывается функция $y = \sqrt{1 + x^2}$ — это, конечно, «выражение, составленное из переменной величины и постоянных величин». (Скажем, кстати, что воззрения Бернулли во многом смыкаются с воззрениями современных «конструктивистов»: надо лишь расплывчатый термин «выражение» заменить термином «алгоритм», которому можно придать совершенно точный смысл, а по сути дела, алгоритм — это точно сформулированное предписание о том, что надлежит сделать с числом x для получения y.)

Функции одной переменной можно изображать графиками. Для этого ось x направим, как обычно, горизонтально слева направо, ось y — вертикально снизу вверх. Пересечение осей обозначим буквой O. Выберем масштаб по оси x и y. Теперь, выбрав число x, отложим его на оси Ox, а затем на перпендикуляре, проведенном через полученную точку, отложим отрезок длины $y = f(x)$. Так получается график функции $y = f(x)$. В частности, графиком функции $y = \sqrt{1 + x^2}$ будет гипербола (см. рис. 40).



А теперь давайте определим и затем изобразим несколько важнейших функций одной переменной.

Самая простая из функций — *постоянная*:

$$y \equiv c.$$

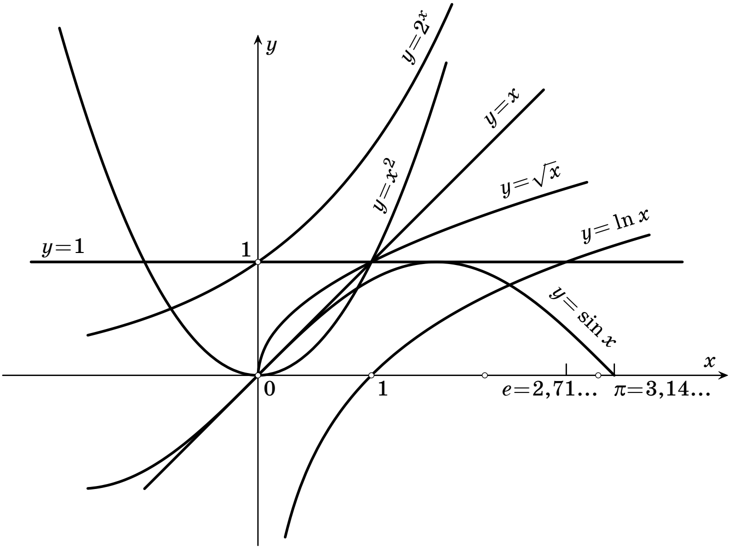
Каждому действительному числу x эта функция сопоставляет одно и то же число c. На рис. 41 изображена функция $f(x) \equiv 1$ при частном значении c, равном единице.

Следующие по сложности функции — *линейные*:

$$y = bx.$$

Здесь b — постоянное число, а x — любое действительное число.

Такие функции изображаются прямыми, проходящими через начало координат и не совпадающими с осью Oy. Пусть, скажем, $b = 2$. Тогда получится функция $y = 2x$. Если взять $x = 1$, то y окажется равным двум, если взять $x = 0{,}25$, то y будет $0{,}5$ и т. д. — по каждому x совершенно определенным образом находится y, а именно, x умножается на 2. На рис. 41 изображена самая



простая из линейных функций — $y = x$. Ее график является биссектрисой угла, образованного осями Ox и Oy.

Из постоянной и линейной функций образуются функции вида

$$y = bx + c.$$

В школе именно такие функции называют линейными, но нам кажется, что их естественней называть аффинными. Аффинные функции в дальнейшем у нас будут играть важнейшую роль.

Затем идут *квадратичные функции*

$$y = ax^2.$$

Такие функции изображаются параболами, проходящими через начало координат. Пусть, скажем, $a = 0{,}5$. Тогда, если взять $x = 1$, то $y = 0{,}5$, если $x = 2$, то $y = 2$, если $x = 10$, то $y = 50$ и т. д., снова по каждому x совершенно определенным образом находится y, а именно, x умножается на себя, а затем на $0{,}5$. На нашем рисунке изображена функция

$$y = x^2 \quad (a = 1).$$

Из постоянной, линейной и квадратичной функций образуются *квадратные трехчлены* $y = ax^2 + bx + c$.

Потом можно рассмотреть *степенные функции*

$$y = Ax^n.$$

Здесь A — какое-то постоянное действительное, а n — целое число.

Важную роль в анализе играют *показательные функции*

$$y = a^x.$$

Функция $y = 2^x$ изображена на рис. 41.

Очень употребительны функции, обратные к степенным и показательным. На нашем рисунке изображены функции $y = \sqrt{x}$ и $y = \ln x$. Особенность этих функций в том, что они определены не для всех x. Функция $y = \sqrt{x}$ определена только для неотрицательных x, а функция $y = \ln x$ — только для положительных x (см. рис. 41).

Наконец, в школе проходят еще и тригонометрические функции. На рис. 41 изображена функция $y = \sin x$.

Приведем список наиболее употребительных функций, с которыми постоянно придется иметь дело:

$y = Ax^n$ — степенна́я функция; при $n = 0$ — это постоянная функция, при $n = 1$ — линейная, при $n = 2$ — квадратичная;

$y = \sqrt[n]{x}$ — функция «корень n-й степени из x»;

$y = a^x$ — показательная функция;

$y = \log_a x$ — логарифмическая функция;

тригонометрические функции: $y = \sin x$ — синус, $y = \cos x$ — косинус, $y = \dfrac{\sin x}{\cos x} = \operatorname{tg} x$ — тангенс, $y = \dfrac{\cos x}{\sin x} = \operatorname{ctg} x$ — котангенс.

Из описанных функций можно составить разнообразные выражения, которые также все окажутся функциями одной переменной. Приведем примеры:

$$y = \sqrt{a^2 + x^2}, \quad y = x\sqrt{1 - x^2}, \quad y = \frac{ax}{2} - \frac{\pi x^3}{3},$$

$$y = \frac{(a + x)^2}{x}, \quad y = \sqrt{b^2 + x^2 - 2bx\cos\alpha}.$$

Конечно, этот список можно продолжить. Мы привели здесь лишь те функции, с которыми нам придется иметь дело, когда нужно будет решать экстремальные задачи. Но, разумеется, можно придумать и что-нибудь пострашнее, например,

$$y = \sin\left(2^{\sqrt[5]{\log_7(1+\operatorname{tg}(x+1))}}\right).$$

Здесь закодировано следующее предписание: взять число, прибавить к нему единицу, взять от суммы тангенс, ко всему прибавить единицу, из полученной суммы взять логарифм по основанию семь, далее следует извлечь из полученного числа корень пятой степени, возвести двойку в степень, равную получившемуся числу и, наконец, найти синус этой степени двойки.

Такого рода «выражениями» («предписаниями», «алгоритмами»), только чуть менее «страшными», в основном и будет исчерпано содержание понятия «функция одной переменной», с которым нам придется сталкиваться.

С функциями одной переменной вы встречались в школе. Поэтому выше о многом (и важном) было сказано вскользь, в расчете на то, что вы можете заглянуть в учебник, спросить учителя или додумать кое-что сами. Но функциями одной переменной нам дальше не обойтись. Нам необходимо будет «работать» с функциями двух, трех, четырех, ста переменных (и — пока по секрету — с функциями бесконечного числа переменных). Но не пугайтесь! Все это не так уж трудно.

Сначала сделаем лишь один шаг вперед и поговорим о функциях *двух переменных*.

Если спросить вас, когда впервые вы начали «работать» с функциями двух переменных, многие придут в недоумение. Функции двух переменных в школе не изучают — как же можно с ними «работать»? Но в действительности все мы знакомимся с функциями двух переменных с незапамятных времен. Именно — с незапамятных! Никто из вас и не вспомнит тот день, когда впервые ваш отец, или мать, или пришедший в гости знакомый задал вам вопрос вроде: «Вот яблоко, а вот еще, — сколько вместе?». Это было в самом раннем детстве, когда едва лишь освоена речь и еще далеко до знания букв. Тот момент, когда вы ответили на поставленный вопрос, вы не можете помнить. Но именно тогда, когда, прибавив к одному яблоку другое, вы получили «два» яблока, вы впервые столкнулись с функцией двух переменных, самой древней из них, самой известной — *сложением*.

Вот она:

$$z = x + y.$$

Если взять любую пару чисел x и y, то этой паре функция «сложение» сопоставит число z, равное сумме x и y. Паре $(1, 1)$ будет сопоставлено число 2, паре $(7, -3)$ — число 4 и т. д.

На первых порах вы осваивали сложение натуральных чисел, затем — целых; наконец, вас учили складывать любые числа. В итоге, *любой паре* (x, y), где x и y — действительные числа, вы сможете сопоставить их сумму, иначе говоря, операция «сложение» определена на *любой* паре (x, y), где x и y — действительные числа.

Выходит, что с функциями двух переменных вы столкнулись гораздо раньше, чем с функциями одной переменной.

После сложения вас учили вычитать, умножать и делить. Вычитание, умножение и деление — все это функции двух переменных. Причем вычитание и умножение определены снова на любой паре (x, y), где x и y — действительные числа, а деление определено только на тех парах (x, y), для которых $y \neq 0$.

И снова под функцией двух переменных мы будем понимать (вслед за Бернулли) *выражение, составленное из переменных величин x и y и постоянных величин*.

Функции двух переменных $z = f(x, y)$ тоже можно изобразить графически. Для этого изобразим плоскость xOy и проведём ось Oz перпендикулярно этой плоскости. Выберем масштаб на осях Ox, Oy и Oz. Теперь, если нам заданы числа x и y, мы отложим их на плоскости xOy, затем вычислим значение $f(x, y)$ и отложим на прямой, параллельной оси Oz и проходящей через точку (x, y), отрезок длины $f(x, y)$.

Давайте теперь определим и изобразим несколько важнейших функций двух переменных.

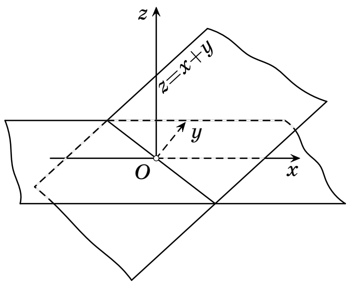
И снова — самая простейшая функция — *постоянная*:

$$z \equiv c.$$

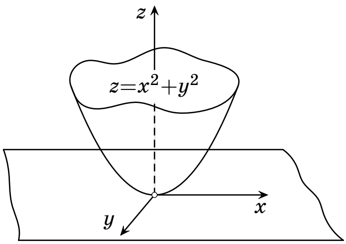
Каждой паре (x, y), где x и y — действительные числа, эта функция сопоставляет число c.

Следующие по сложности функции — *линейные*:

$$z = ax + by.$$



Здесь a и b — постоянные числа, а x и y — любые действительные числа. Для того чтобы получить z по паре (x, y), надлежит x умножить на a, y умножить на b и получившиеся числа сложить. Пусть, скажем, $a = 2$, $b = 0{,}5$. Тогда паре $(1, 0)$ соответствует $z = 2$, паре $(1, 1)$ соответствует $z = 2{,}5$, паре $(4, -16)$ соответствует $z = 0$. На рис. 42 изображена линейная функция $z = x + y$. Вообще же линейные функции изображаются



плоскостями, проходящими через начало координат и не перпендикулярными плоскости xOy.

Из постоянной и линейных функций образуются функции вида

$$z = ax + by + c,$$

которые мы также будем называть *аффинными*. Они далее будут играть для нас самую важную роль.

Дальше идут квадратичные функции $z = Ax^2 + 2Bxy + Cy^2$ и функции вида

$$z = Ax^2 + 2Bxy + Cy^2 + ax + by + c.$$

На рис. 43 изображена функция $z = x^2 + y^2$. Приведем еще несколько функций двух переменных, которые встретятся у нас далее при решении задач:

$$z = Ax^2(y - Bx), \quad z = Axy, \quad z = \frac{x^2}{a^2} + \frac{y^2}{b^2},$$

$$z = (x-c)^2 + (y-d)^2, \quad z = \sqrt{(x-c)^2 + (y-d)^2}.$$

И еще несколько слов о функциях двух переменных. Одно обстоятельство позволяет нам наглядно представлять их характерные особенности, ибо график одной замечательной функции двух переменных всегда у нас перед глазами и «под ногами».

Вот вы стоите на земле. Ваше положение в пространстве можно задать тройкой чисел (φ, ϑ, h), где (φ, ϑ) — географические координаты — широта и долгота, а h — *высота над уровнем моря*. Таким образом, $h = h(\varphi, \vartheta)$, т. е. h есть функция (φ, ϑ). (Вспомним ту общую точку зрения на функцию, которую мы поставили эпиграфом к данному рассказу.) И все, что мы видим — холмы, ложбины, овраги, горы, глади озер — все это есть «график» этой функции.

У неё имеется множество примечательных точек.

Прежде всего, конечно, это — *вершины*, поднявшись на которые мы испытываем радость победы, восторг преодоления. Во время путешествия по горам мы ищем *перевалы*, чтобы из одной долины перейти в другую. Исследуя морское дно, мы стремимся отыскать самую глубокую точку моря. Этими же самыми точками — вершинами, перевалами, ложбинами — нам и надлежит заниматься дальше.

Расстанемся на время с функциями двух переменных и сделаем еще один шаг. Спросим себя: что такое функция трёх переменных? Ну конечно же, это *выражение, составленное из переменных x, y и z и постоянных величин.*

Примеры? Сколько угодно! Сумма $u = x + y + z$, произведение $u = x \cdot y \cdot z$, постоянная функция $u \equiv d$, линейная функция $u = ax+by+cz$, квадратичная функция $u = Ax^2+2Bxy+Cy^2+2Dyz+ + 2Gxz+Fz^2$, $u = \sqrt{x^2+y^2+z^2}$, $u = 2^x + \log_3 y + \sin z$ и т. д., и т. п.

А что такое функция четырех переменных? Ну конечно же, — *выражение, составленное из переменных x, y, z, u и постоянных величин.*

А что такое функция двадцати шести переменных? Это, конечно, — выражение, составленное из переменных a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z и постоянных величин.

А из двадцати семи? Это, конечно, — выражение, составленное из переменных... Но ведь букв не хватит! Неужели изучать лишь функции не более 26 переменных? Конечно, можно еще добавить греческий алфавит и еще какие-то другие алфавиты. Но и всего этого будет мало. В современных экономических задачах — тысячи переменных! Что же делать?

Выход очень прост. Переменные можно обозначать не разными буквами, а одной буквой x, но с индексами: $x_1, x_2, \ldots, x_n$. Постоянные же можно обозначить буквами начала латинского алфавита, и снова с разными индексами.

Таким образом, функции n переменных — *это выражение, составленное из переменных x_1, x_2, ..., x_n и постоянных величин.*

Простейшая функция n переменных — *постоянная*: $y \equiv c$. Каждому набору из n чисел $(x_1, \ldots, x_n)$ она сопоставляет единственное число c.

Важнейший класс образуют *линейные функции*

$$y = a_1x_1 + \ldots + a_nx_n.$$

Из постоянных и линейных функций образуются функции вида

$$y = a_1x_1 + \ldots + a_nx_n + c.$$

Их называют *аффинными*.

Такого рода функциями мы будем аппроксимировать более сложные функции.

И теперь уже, быть может, настало время назвать вещи своими именами: мы с вами уже много раз сталкивались с функциями многих переменных. Это было, например, в пятом рассказе, когда речь шла о средних. Мы рассматривали там функции

$$y = x_1 \ldots x_n, \quad y = \sqrt[n]{x_1 \ldots x_n},$$

$$y = \sqrt{x_1^2 + \ldots + x_n^2}, \quad (x_1^p + \ldots + x_n^p)^{1/p}, \quad x_i \geqslant 0,$$

и т. д. Функции многих переменных встречались и в рассказе о брахистохроне:

$$T_n = \sqrt{y_1^2 + x_1^2}/\sqrt{2gy_1} + \ldots + \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2}/\sqrt{2gy_2} + \ldots$$
$$\ldots + \sqrt{(y_n - y_{n-1})^2 + (x_n - x_{n-1})^2}/\sqrt{2gy_n},$$

и в рассказе о задаче Ньютона. И там уже возникли функции еще более поразительные — «функции от функции», например от линии, т. е. в конечном счете — функции бесконечного числа переменных.

Наша цель в этой части книги — дать фрагмент общей теории экстремальных задач. Пока что мы объяснили, что такое функции многих переменных. В следующем рассказе мы объясним, как ставятся задачи на максимум и минимум функций многих переменных с ограничениями.

Рассказ десятый

ЧТО ТАКОЕ ЭКСТРЕМАЛЬНАЯ ЗАДАЧА?

В первой части этой книги мы перерешали множество задач на максимум и минимум. Некоторые из них были поставлены очень давно, сотни лет назад, и в их исследованиях приняли участие величайшие математики прошлого — Евклид, Архимед, Кеплер, Ферма, Бернулли, Лейбниц, Ньютон... Эти исследования были совсем не похожи друг на друга, а цель иногда достигалась после долгих блужданий.

Эпиграфом к этой части поставлены слова Даламбера. Вдумаемся в них. Спросим себя: возможно ли изыскать способ решения всех задач (и даже тех, о которых рассказывалось в первой части) одним и притом простым способом?

То, что не может быть единого простого рецепта решения всех на свете проблем, ясно каждому. (Даламбер, кстати, говорил лишь о задачах динамики.) Может показаться сомнительной и возможность такого единого приема, который позволил бы решить хотя бы все задачи, обсуждавшиеся нами в первой части. Но тем не менее такой прием имеется, и мы изложим его здесь, во второй части этой книги. С помощью этого (одного и притом простого) приема все задачи первой части будут решены единообразно, стандартно, если хотите — рутинно.

Но сначала нужно позаботиться о единой форме записи условий задач и об общем языке, на котором можно было бы рассуждать о задачах столь разного содержания. Этому посвящен наш рассказ.

Сначала давайте еще раз уточним значения ключевых слов: «максимум», «минимум», «экстремум», «оптимум».

Напомним о двух планиметрических задачах из первой части. Об одной рассказывалось неоднократно, начиная с первого рассказа, о другой говорилось в рассказе пятом.

Задача Герона. *Даны две точки по одну сторону от прямой. Требуется на прямой найти такую точку, чтобы сумма расстояний от нее до двух данных точек была минимальна* (рис. 1).

Планиметрическая задача Кеплера. *Вписать в круг единичного радиуса прямоугольник наибольшей площади.*

Эти задачи отличаются друг от друга тем, что в первой надо найти минимум, а во второй — максимум.

Напомним: слова maximum и minimum — латинские. Они значат «наибольшее» и «наименьшее». Еще два слова латинского происхождения часто употребляют, когда говорят про задачи на максимум и минимум. Термин «экстремум» — от латинского extremum, что значит «крайнее» — объединяет понятия максимум и минимум (этот термин предложил использовать французский ученый Дюбуа-Реймон). В последние годы вошло во всеобщее употребление прилагательное «оптимальный». Оно происходит от латинского optimus, означающего «наилучший», «совершенный».

Теорию задач на отыскание наибольших и наименьших величин называют или *теорией экстремальных задач*, или *теорией оптимизации*. А если проблема состоит в наилучшем воздействии на какие-то процессы и явления, которыми человек может в определенных пределах управлять, то ее относят к разделу теории экстремальных задач, называемому *оптимальным управлением*.

Задачи Герона и Кеплера были чуть выше сформулированы словесно, без формул. То же самое относится и ко всем задачам первой части: в задачах геометрического содержания (изопериметрической, штейнеровской и др.) использовалась геометрическая терминология, в механических (о брахистохроне, ньютоновской) — механическая, в алгебраических (тартальевской, о неравенствах и др.) — алгебраическая. Никаких формул не было. И все это — в порядке вещей. Экстремальные проблемы, возникающие в математике, естествознании или в практических делах, первоначально ставятся без формул, в терминах той области, в которой они возникли. Для того чтобы можно было воспользоваться общей теорией, необходимо осуществить п е р е в о д постановок задач со специфического языка на язык математики. Такой перевод называется *формализацией*.

Проиллюстрируем, как осуществляется формализация на примерах. Начнем с задачи Герона.

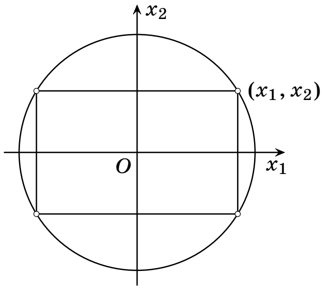
Направим ось Ox по заданной прямой, а ось Oy проведем перпендикулярно оси Ox через точку A (рис. 1). Пусть координаты точек A и B оказались такими: $A = (0, a)$, $B = (d, b)$. Возьмем на оси Ox точку D с координатой $(x, 0)$. Тогда сумма расстояний от A до D и от D до B будет равна $\sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2}$.

Мы пришли к такой задаче: *найти наименьшее значение функции*

$$f(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2}$$

по всем x.

Приведенная нами формализация выглядит очень естественно и как бы напрашивается сама собой. Но, вообще говоря, задача может иметь множество формализаций. Покажем это на примере планиметрической задачи Кеплера.



Направим оси Ox_1 и Ox_2 параллельно сторонам прямоугольника. Единичная окружность описывается уравнением $x_1^2 + x_2^2 = 1$. Пусть (x_1, x_2) — координаты той вершины прямоугольника, которая лежит в первом квадранте (см. рис. 44). Тогда площадь прямоугольника равна $4x_1x_2$. Получилась такая формализация: найти наибольшее значение функции двух переменных

$$f_0(x_1, x_2) = 4x_1x_2$$

при условиях

$$f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0, \quad f_2(x_1, x_2) = x_1 \geqslant 0,$$
$$f_3(x_1, x_2) = x_2 \geqslant 0.$$

Заметим теперь, что неравенства $x_1 \geqslant 0$ и $x_2 \geqslant 0$ можно отбросить. Вы легко сообразите, что задача о нахождении наибольшего значения функции

$$f_0(x_1, x_2) = 4x_1x_2$$

при условии

$$f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$$

также является формализацией планиметрической задачи также является Кеплера.

Но теперь из уравнения $f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0$ можно выразить x_2 через x_1 и подставить в выражение для f_0. Тогда (если x_1 обозначить через x) получится еще одна формализация: *найти наибольшее значение функции* $\varphi(x) = 4x\sqrt{1 - x^2}$ *при условии* $0 \leqslant x \leqslant 1$ *или при условии* $|x| \leqslant 1$ (это условие накладывается областью определения функции и смыслом переменной x).

Мы видим, что одну и ту же задачу можно формализовать по-разному. От того, насколько удачно формализована задача, часто зависит и успех ее решения. Формализация — это искусство. Ему нужно учиться, лучше всего — решая практические задачи.

Задача Кеплера, о которой мы здесь говорили, — планиметрическая. В пространстве можно поставить подобные задачи по-разному. Одна из постановок была обсуждена нами в первой части (вписать в единичный шар цилиндр наибольшего объема). Но возможна и другая постановка: *вписать в единичный шар прямоугольный параллелепипед наибольшего объема*. Эта постановка приводит к такой формализации: найти наибольшее значение функции трех переменных

$$f_0(x_1, x_2, x_3) = 8x_1x_2x_3$$

при условии, что

$$f_1(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 1 = 0.$$

(Частный случай этой задачи, когда $x_1 = x_2$, рассматривал, как мы помним, Кеплер.)

К той же формализации (правда, без восьмерки) приводит такая задача: *найти наибольшее значение произведения трех чисел при условии, что сумма квадратов их равна заданному числу*. Но теперь возможно и дальнейшее обобщение: вместо трех чисел можно взять и пять, и десять, и вообще любое количество. Формализация этой задачи будет такова: найти наибольшее значение функции n переменных

$$f_0(x_1, \ldots, x_n) = x_1 \ldots x_n$$

при условии, что

$$f_1(x_1, \ldots, x_n) = x_1^2 + \ldots + x_n^2 - 1 = 0.$$

Теперь попытаемся уже достаточно определенно описать все элементы точно формализованной экстремальной задачи. Там обязательно должна присутствовать некоторая функция (скажем, n переменных), наибольшее или наименьшее значение которой надо найти (максимизируемая или минимизируемая функция).

Далее в формализованной задаче должно быть *ограничение*, задаваемое некоторым числом равенств или неравенств с теми же неизвестными.

Приведем несколько минимизируемых или максимизируемых функций и функций, задающих ограничения, из тех, что уже встретились нам или встретятся далее.

$f_1(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2}$ (задача Герона);

$f_2(x) = \dfrac{\sqrt{a^2 + x^2}}{v_1} + \dfrac{\sqrt{b^2 + (d - x)^2}}{v_2}$ (задача о преломлении света);

$f_3(x) = \dfrac{H}{b} x(b - x)$ (задача Евклида);

$g_1(x_1, x_2) = 4x_1x_2,$
$g_2(x_1, x_2) = x_1^2 + x_2^2 - 1$ (планиметрическая задача Кеплера);

$h_1(x_1, x_2, x_3) = 8x_1x_2x_3,$
$h_2(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^3 - 1$ (задача о параллелепипеде, вписанном в шар);

$F_0(x_1, \ldots, x_n) = x_1 \ldots x_n,$
$F_1(x_1, \ldots, x_n) = x_1^2 + \ldots + x_n^2 - 1.$

Здесь f_1, f_2, f_3 — функции одного переменного, g_1, g_2 — двух; h_1, h_2 — трех, F_0, F_1 — n переменных.

Таким образом, *формализовать экстремальную задачу — это значит точно описать минимизируемую или максимизируемую функцию* (обозначим её f_0) *и ограничение* (обозначим его C). Ограничение задается обычно равенствами и неравенствами.

Для формализованной задачи: «*найти минимум* (*максимум*) *функции* $f_0(x)$ *при условии, что* x из C» будем употреблять сокращенную запись *)

$$f_0(x) \to \min\ (\max) \text{ по } x \text{ из } C. \tag{з}$$

Точки из C называются *допустимыми*; если ограничений нет, то (з) называют *задачей без ограничений*.

Например, для одной из формализаций планиметрической задачи Кеплера сокращенная запись примет вид

$$f_0(x_1, x_2) = 4x_1x_2 \to \max, \quad f_1(x_1, x_2) = x_1^2 + x_2^2 - 1 = 0, \tag{з$_1$}$$

а для задачи Герона будет просто

$$f_0(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2} \to \min. \tag{з$_2$}$$

Задача (з$_1$) — *с ограничением типа равенств*, (з$_2$) — *задача без ограничений*.

Допустимая точка $\widehat{x}$ называется *абсолютным минимумом* в задаче (з), если $f(x) \geqslant f(\widehat{x})$ для любого x из C (соответственно *максимумом*, если $f(x) \leqslant f(\widehat{x})$ для любого x из C). Абсолютный минимум (максимум) задачи будем называть *решением задачи*. Найти решение и есть наша цель.

Но для того чтобы достичь ее, мы будем использовать методы нахождения так называемых *локальных экстремумов*.

---

*) Буквы «з» не следует бояться — это русская буква, первая буква в слове «задача». Для краткости набор переменных $(x_1, \ldots, x_n)$ обозначается буквой x. Далее мы говорим иногда «точка x». При этом, если $x = (x_1, \ldots, x_n)$ и a — число, то $ax = (ax_1, \ldots, ax_n)$, и если $x' = (x'_1, \ldots, x'_n)$, то $x + x' = (x_1 + x'_1, \ldots, x_n + x'_n)$.

Взобравшись на вершину холма или маленькой горки, расположенной на ровной местности, мы оказываемся в наивысшей точке этой местности. Но это не значит, что мы нашли решение задачи о самой высокой точке над уровнем моря. Эту последнюю задачу «решили» лишь те люди, которые побывали на Эвересте.

Такова разница между абсолютным и локальным экстремумами.

Дадим точное определение последнего понятия. Будем говорить, что точка $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ доставляет локальный минимум (максимум) для задачи (з), если можно указать такое число $\varepsilon > 0$, что для всех точек $x = (x_1, \ldots, x_n)$ из C, для которых

$$\sqrt{(x_1 - \widehat{x}_1)^2 + \ldots + (x_n - \widehat{x}_n)^2} < \varepsilon,$$

выполнено неравенство

$$f_0(x) \geqslant f_0(\widehat{x}) \quad (f_0(x) \leqslant f_0(\widehat{x})).$$

(Иначе говоря, «вблизи» от точки $\widehat{x}$ значение функции в допустимой точке не больше (не меньше) $f_0(\widehat{x})$.)

Прежде чем закончить этот рассказ, формализуем еще одну задачу, а именно транспортную, о которой мы мимоходом упомянули в первом рассказе. Напомним, что в транспортной задаче нужно составить план перевозок, для того чтобы доставить некоторый продукт с баз в магазины и чтобы стоимость перевозок была минимальной. Обозначим через a_i количество единиц груза на i-й базе. При этом число баз пусть будет равно m. Потребность j-го магазина в единицах перевозимого груза пусть исчисляется величиной b_j, $j = 1, \ldots, n$. Наконец, через c_{ij} выразим стоимость перевозки единицы груза с i-й базы в j-й магазин. Составить план перевозок — это значит указать числа x_{ij}, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, единиц груза, которые следует перевезти с i-й базы в j-й магазин. Таким образом, количество груза, вывозимого с i-й базы, будет равно $x_{i1} + \ldots + x_{in}$, и это число не должно превосходить a_i (никакая база не может дать больше, чем у нее есть), а количество продукта, привозимого в j-й магазин, равно $x_{1j} + \ldots + x_{mj}$, и это число должно быть равно в точности потребности магазина, т. е. b_j. Таким образом, транспортная задача допускает следующую формализацию:

$$c_{11}x_{11} + \ldots + c_{1n}x_{1n} + \ldots + c_{m1}x_{m1} + \ldots + c_{mn}x_{mn} \to \min,$$
$$x_{i1} + \ldots + x_{in} \leqslant a_i, \quad i = 1, \ldots, m,$$
$$x_{1j} + \ldots + x_{nj} = b_i, \quad j = 1, \ldots, n,$$
$$x_{ij} \geqslant 0, \quad 1 \leqslant i \leqslant m, \quad 1 \leqslant j \leqslant n.$$

Обратим внимание на то, что и минимизируемая функция, и ограничения в транспортной задаче задаются с помощью *линейных функций*. Раздел теории экстремальных задач, где изучают экстремумы линейных функций при линейных ограничениях, называется *линейным программированием*. К этому разделу относится и транспортная задача.

Следующий наш рассказ — об экстремумах функций одной переменной. Сейчас эту тему проходят в школе, но мы не побоимся некоторых повторений.

Рассказ одиннадцатый

ЭКСТРЕМУМЫ ФУНКЦИЙ ОДНОЙ ПЕРЕМЕННОЙ

Когда величина является максимальной или минимальной, в этот момент она не течет ни вперед, ни назад.

И. Ньютон

Этот и следующий рассказы построены одинаково. Они состоят из двух частей. В первой части излагается сам метод решения задач без доказательств, хотя и с некоторыми объяснениями и комментариями. Во второй части проводятся точные определения и некоторые доказательства. Для освоения первой части необходимо владеть лишь следующими понятиями: «предел», «непрерывная функция» и «производная».

1. Здесь будет рассказано о методе отыскания решения таких задач на экстремум функций одного переменного:

$$f_0(x) \to \min\ (\max), \quad a \leqslant x \leqslant b. \tag{з}$$

В (з) a и b могут принимать и бесконечные значения. Таким образом, речь будет идти об экстремумах функции f_0 на конечном отрезке, на луче или на совокупности всех вещественных чисел.

Примеры.

$$f_0(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d - x)^2} \to \min, \tag{з$_1$}$$

$$f_0(x) = x\sqrt{1 - x^2}, \quad -1 \leqslant x \leqslant 1. \tag{з$_2$}$$

Напомним, что (з$_1$) — это формализация задачи Герона, (з$_2$) — формализация планиметрической задачи Кеплера (обе задачи были формализованы в рассказе десятом).

Заметим, что не всякая задача имеет решение. Например, мы уже рассматривали такую задачу без ограничений:

$$f_0(x) = -\frac{1}{1 + x^2} \to \max. \tag{з$_3$}$$

Функция $f_0(x) \leqslant 0$, и нет такой точки $\bar{x}$, где $f_0(\bar{x}) = 0$.

С другой стороны, если взять точки $x_n = n$, $n = 1, 2, \ldots$, то $f_0(x_n) \to 0$. Отсюда следует, что максимума в (з$_3$) не существует, т. е. нельзя указать такую точку $\widehat{x}$, что $f_0(x) \leqslant f_0(\widehat{x})$ для всех x.

Итак, максимумы и минимумы существуют не всегда. Однако имеется замечательная теорема Вейерштрасса, которая дает в огромном числе случаев гарантию существования решения.

Т е о р е м а В е й е р ш т р а с с а. *Пусть* $f_0(x)$ — *непрерывная функция на конечном отрезке* $[a, b]$. *Тогда решения обеих задач*

$$(\text{з}_{\min}) \qquad f_0(x) \to \min, \quad a \leqslant x \leqslant b,$$

и

$$(\text{з}_{\max}) \qquad f_0(x) \to \max, \quad a \leqslant x \leqslant b,$$

существуют.

Из этой теоремы немедленно следует, что решение задачи (з_2) существует. А про (з_1) этого сказать пока нельзя, ибо там функция рассматривается на всей прямой, а не на конечном отрезке.

Из теоремы Вейерштрасса выводится такое следствие, которое позволит, в частности, доказать существование решения и в (з_1).

С л е д с т в и е. *Пусть* f_0 *непрерывна на всей прямой. Тогда, если* $\lim\limits_{x\to\infty} f_0(x) = \lim\limits_{x\to-\infty} f_0(x) = \infty$, *решение задачи без ограничений*

$$f_0(x) \to \min$$

существует.

У нас еще встретятся случаи, когда f_0 непрерывна на луче вида $a \leqslant x < \infty$ или $a < x < \infty$. Следовательно, *если в первом случае* $\lim\limits_{x\to\infty} f_0(x) = \infty$, *а во втором* $\lim\limits_{x\to a} f_0(x) = \lim\limits_{x\to\infty} f_0(x) = \infty$, *то функция* f_0 *достигает своего минимума на соответствующем луче.*

Для отыскания решения задачи (з) будем использовать прием, впервые примененный Ферма.

Но сначала напомним одно определение, о котором говорилось в предыдущем рассказе. Пусть функция f_0 определена на отрезке $a \leqslant x \leqslant b$ и $\widehat{x}$ — точка из этого отрезка. Говорят, что точка $\widehat{x}$ доставляет локальный минимум (максимум) в задаче (з), если можно указать такое ε, что для всех точек x из отрезка $[a, b]$, для которых $|x - \widehat{x}| < \varepsilon$, выполнено неравенство $f_0(x) \geqslant f_0(\widehat{x})$ ($f_0(x) \leqslant f_0(\widehat{x})$).

Иногда мы говорим проще: x доставляет локальный экстремум функции f_0.

Имеет место следующая теорема.

Т е о р е м а Ф е р м а. *Пусть функция* f_0 *является дифференцируемой в точке* $\widehat{x}$. *Тогда, если точка* $\widehat{x}$ *доставляет локальный экстремум (минимум или максимум) этой функции* f_0, *то* $f_0'(\widehat{x}) = 0$.

Точки, для которых $f_0'(x) = 0$, называются *стационарными*. Стационарные точки совместно с концевыми точками называются *критическими*.

Соотношение $f_0'(\widehat{x}) = 0$ является лишь *необходимым* условием экстремума. Например, для функции $f_0(x) = x^3$ точка $\widehat{x} = 0$ является стационарной, но ни локального максимума, ни локального минимума не доставляет.

Теорема Ферма позволяет дать следующее правило поиска решений одномерных задач. Разобьем его на 4 этапа.

Этап 1 — формализация задачи. Требуется привести (разумеется, если это возможно) стоящую перед вами задачу к виду

$$f_0(x) \to \min(\max), \quad a \leqslant x \leqslant b. \tag{з}$$

Этап 2 состоит в выписывании необходимого условия $f_0'(x) = 0$.

Этап 3 состоит в нахождении всех стационарных точек.

Этап 4 состоит в переборе всех критических значений функции f_0 и выборе минимального (максимального) среди них.

Из теорем Вейерштрасса и Ферма следует, что если функция f_0 удовлетворяет на $[a, b]$ теореме Вейерштрасса (или следствиям из нее) и, кроме того, если она дифференцируема во внутренних точках x отрезка $[a, b]$ (когда $a < x < b$), то описанное правило приведет к решению задачи.

Выделим отдельно следующий факт, которым в основном и будем пользоваться: *если отрезок* $[a, b]$ *конечен, функция* f_0 *непрерывна на* $[a, b]$ *и дифференцируема во внутренних точках* x, $a < x < b$, *то решение находится среди критических точек.*

Таким образом, для применения описанного правила требуется умение дифференцировать. Для облегчения этой процедуры приведем таблицу производных основных функций, несколько часто употребляемых формул, а также напомним важнейшее правило нахождения производной сложной функции.

Таблица производных

	$f(x)$	$f'(x)$
1	$x^a \ (a \neq 0, x > 0)$	ax^{a-1}
2	$a^x \ (a \neq 1, a > 0)$	$a^x \ln a$
3	$\log_a x \ (a \neq 1, x > 0)$	$\dfrac{1}{x \ln a}$
4	$\ln x$	$1/x$
5	$\sin x$	$\cos x$
6	$\cos x$	$-\sin x$

Кроме этой таблицы, полезно помнить следующие формулы:

7. $(f+g)' = f' + g'$, $(fg)' = f'g + fg'$, $\left(\dfrac{f}{g}\right)' = \dfrac{f'g - g'f}{g^2}$.

И еще одно. Очень часто функции, с которыми нам придется сталкиваться, имеют вид $h(x) = f(g(x))$. Нужно запомнить и научиться пользоваться следующей формулой для производной сложной функции:

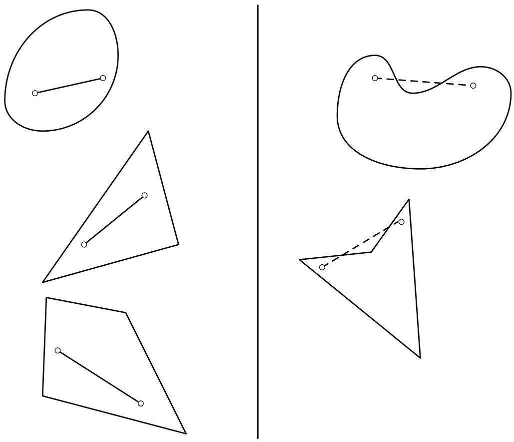
8. $h'(x) = f'(g(x))\,g'(x)$.

Пример. $h(x) = \sqrt{a^2 + x^2}$. Здесь $h(x) = f(g(x))$, где $f(u) = \sqrt{u} = u^{1/2}$; $g(x) = a^2 + x^2$. Пользуясь формулой 1 таблицы и формулами 7 и 8, получаем

$$h'(x) = \frac{1}{2\sqrt{a^2+x^2}} \cdot (a^2 + x^2)' = \frac{2x}{2\sqrt{a^2+x^2}} = \frac{x}{\sqrt{a^2+x^2}}.$$

В заключение этого пункта скажем несколько слов о *выпуклых функциях*. Их значение в теории экстремальных задач велико, и нам придется не раз затрагивать эту тему. Сейчас дадим определение выпуклых функций одной переменной.

С самим понятием выпуклости мы сталкиваемся еще в школе. Напомним, что *фигура называется выпуклой*, если вместе с любыми двумя своими точками она содержит весь отрезок, соединяющий эти точки. Любой треугольник как часть плоскости — выпуклая фигура, среди четырехугольников встречаются и невыпуклые (рис. 45).



Можно дать три равносильных определения выпуклой функции. Приведем их. Функция $y = f(x)$ называется *выпуклой*, если для любой хорды, соединяющей две точки графика этой функции, ее график в промежуточных точках лежит ниже этой хорды; или: множество точек плоскости, лежащее выше графика функции $y = f(x)$, является выпуклым; или: для любых чисел x_1 и x_2 и любого α, $0 \leqslant \alpha \leqslant 1$, имеет место следующее неравенство (*неравенство Иенсена*):

$$f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha f(x_1) + (1-\alpha)f(x_2).$$

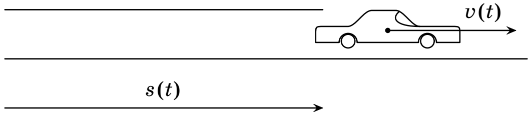
Примеры выпуклых функций доставляют прежде всего линейные функции, функции вида $y = bx + c$ (аффинные) и квадратные трехчлены $y = ax^2 + bx + c$, у которых $a > 0$. Среди функций $y = |x|^p$ выпуклы лишь те, у которых $p \geqslant 1$. Функция $\sqrt{h^2 + x^2}$ выпукла при любых h. Не все выпуклые функции всюду дифференцируемы. Например, функция $y = |x|$ не дифференцируема в нуле. Но если выпуклая функция дифференцируема, то ее производная является возрастающей функцией.

2. В первой части этого рассказа было описано правило решения задач. Это правило нетрудно запомнить, и сразу можно решать задачи (что мы и будем делать в тринадцатом рассказе). Но многим, наверное, захочется все-таки понять, откуда взялось это правило. Об этом и будет рассказано здесь, причем дважды, и вот почему.

Те, кто увлекаются чтением научной и научно-популярной литературы, разделяются, как мне кажется, на две группы. Одну группу образуют читатели (и их большинство), которые стремятся постичь лишь основные идеи. Они удовлетворяются не вполне строгим, но выразительным изложением и не бывают в претензии, если заметят, что при изложении были опущены некоторые, по впечатлению — несущественные, детали. В первой части этого пункта и аналогичного пункта следующего рассказа я буду ориентироваться на эту группу читателей.

Но не следует забывать и о тех, кого не удовлетворяет лишь описание общих идей, кому захочется разобраться в сути дела, по возможности, до конца. Заключительная часть этого пункта написана в расчете на вторую группу читателей. В этой части мы стремимся быть возможно точными и краткими.

Представьте себе: вы едете по прямолинейному шоссе (рис. 46). В каждый момент времени ваша машина находится на определенном расстоянии от какого-то начального пункта. Таким образом, местоположение машины может быть выражено в каждый заданный момент t одним числом $s(t)$. Получается *функция от времени:* $s(t)$ есть *расстояние* от машины до начального пункта в момент времени t.



Теперь посмотрим на спидометр. Он показывает *скорость*. Скорость машины в момент t обозначим $v(t)$. Из курсов физики и математики мы знаем, что *скорость* — это *производная* пути по времени:

$$v(t) = \frac{ds}{dt} = s'(t).$$

(Говорят, что лорд Кельвин, один из крупнейших физиков XIX века, утверждал нечто обратное. Он говорил примерно так: «Не морочьте мне голову с вашей математикой: *производная — это скорость*»!)

Если скорость в данный момент не равна нулю — пусть для определенности она положительна, как это изображено на рис. 46, — то в следующие моменты мы удалимся от начального пункта на еще большее расстояние, а в предшествующие моменты мы были чуть ближе к начальному пункту. Таким образом, функция расстояния $s(t)$ в данный момент не может иметь ни максимума, ни минимума. И следовательно, *в точке максимума или минимума скорость обязана равняться нулю*: в этот момент мы, по словам Ньютона, не течем «ни вперед, ни назад». Но в этом как раз и состоит теорема Ферма.

Теперь о том же самом скажем поточнее. Прежде всего, необходимо дать точное определение производной. Можно было бы воспользоваться определением, которое проходят в школе, но тогда у нас возникли бы сложности в следующих рассказах, когда речь зайдет о производных функций многих переменных. Поэтому здесь будет дано определение, которое одинаково пригодно и в конечномерном, и (снова пока по секрету) в бесконечномерном случаях.

Итак, что значит: *функция f дифференцируема в данной точке x_0* (или, что то же, *имеет производную* в точке x_0)? Если не прибегать к формулам, то можно сказать так: это означает, что функция $f(x_0 + x) - f(x_0)$ *хорошо аппроксимируется линейной* *функцией*. А если точно, это означает следующее.

Определение. Говорят, что функция $y = f(x)$, определенная на отрезке $[a, b]$, содержащем внутри себя точку x_0 ($a < x_0 < b$), *дифференцируема в точке* x_0 (или, что то же, *имеет производную в точке* x_0), если существует такая линейная функция $y = kx$, что

$$f(x_0 + x) - f(x_0) = kx + r(x),$$

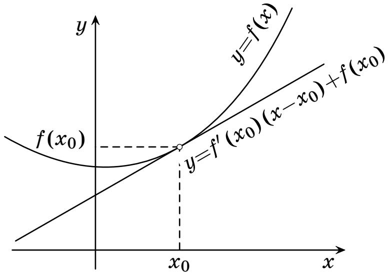
где $\lim\limits_{x\to 0} |r(x)|/|x| = 0$ (или, как иногда говорят, $r(x)/x$ есть величина бесконечно малая).

Из нашего определения немедленно вытекает, что

$$k = \lim_{x\to 0} \frac{f(x_0 + x) - f(x_0)}{x},$$

и, таким образом, число k, участвующее в определении, определяется однозначно. Оно и называется производной f в точке x_0 и обозначается $f'(x_0)$.

Геометрический смысл производной состоит в том, что прямая, являющаяся графиком функции $y = f'(x_0)(x - x_0) + f(x_0)$ (проходящая через точку $(x_0, f(x_0))$ с угловым коэффициентом, равным производной $f'(x_0)$), является касательной к графику функции $y = f(x)$ (рис. 47).



Пример 1. Квадратный трехчлен $y = ax^2 + bx + c$ всюду дифференцируем и его производная в точке x_0 равна $2ax_0 + b$. Проверим это на функции $f(x) = x^2$. Имеем

$$f(x_0 + x) - f(x_0) = (x_0 + x)^2 - x_0^2 = 2x_0x + x^2.$$

При этом $2x_0x$ — линейная функция, $r(x) = x^2$, и значит, $\lim\limits_{x\to 0} |r(x)|/|x| = \lim\limits_{x\to 0} x^2/|x| = 0$, т. е. f дифференцируема в x_0 и $f'(x_0) = 2x_0$.

Из списка, приведенного в п. 1, видно, что элементарные функции a^x, $\sin x$, $\cos x$, $\log_a x$ дифференцируемы всюду, где они определены.

Приведем пример функции, которая в некоторой точке не дифференцируема.

Пример 2. Функция $y = |x|$ не дифференцируема в нуле. Действительно, возьмем любую линейную функцию $y = kx$. Пусть для определенности $k \leqslant 0$. Положим

$$r(x) = f(x) - f(0) - kx = |x| - kx = \begin{cases} x + |k|x, & x \geqslant 0, \\ -x + |k|x, & x \leqslant 0. \end{cases}$$

Значит, $\lim\limits_{x \to 0,\, x \geqslant 0} |r(x)|/|x| = 1 + |k| \neq 0$; функция не дифференцируема. Случай $k \geqslant 0$ аналогичен.

Теорема Ферма. *Пусть функция $f_0(x)$ определена на некотором отрезке $[a, b]$, содержащем внутри себя точку $\widehat{x}$ ($a < \widehat{x} < b$). Пусть при этом она является дифференцируемой в точке $\widehat{x}$. Тогда, если точка $\widehat{x}$ доставляет локальный экстремум (минимум или максимум) этой функции, $f_0'(\widehat{x}) = 0$.*

Точное определение локального экстремума было дано в п. 1.

Доказательство. Предположим, что $f_0'(\widehat{x}) = k \neq 0$, и докажем, что $\widehat{x}$ не является локальным экстремумом. Допустим, $k > 0$. По определению предела из того, что $\lim\limits_{x \to 0} |r(x)|/|x| = 0$ (где $r(x) = f_0(\widehat{x} + x) - f_0(\widehat{x}) - kx$), следует, что найдется $\delta > 0$ такое, что если $|x| < \delta$, то $|r(x)| < \frac{k}{2}|x|$. Но тогда $r(x) \geqslant -\frac{k}{2}x$ для $x > 0$, и следовательно,

$$f_0(\widehat{x} + x) = f_0(\widehat{x}) + kx + r(x) \geqslant f_0(\widehat{x}) + kx - \frac{k}{2}x = f_0(\widehat{x}) + \frac{k}{2}x > f_0(\widehat{x}),$$

и $r(x) \leqslant -\frac{k}{2}x$ для $x < 0$, и следовательно,

$$f_0(\widehat{x} + x) = f_0(\widehat{x}) + kx + r(x) \leqslant f_0(\widehat{x}) + kx - \frac{k}{2}x = f_0(\widehat{x}) + \frac{kx}{2} < f_0(\widehat{x}),$$

т. е. слева от $\widehat{x}$ значения f_0 меньше $f_0(\widehat{x})$, а справа — больше. Это значит, что $\widehat{x}$ не является ни максимумом, ни минимумом. Теорема доказана.

Геометрический смысл теоремы Ферма: *в точке максимума или минимума касательная должна быть горизонтальна*. Подчеркнем еще «вычислительный» смысл экстремума, о котором говорит Кеплер (см. эпиграф к шестому рассказу). Возьмем для примера функции $f_1(x) = x$ и $f_2(x) = x^2$. У первой функции нет экстремума в нуле, а у второй — есть. Если придать приращение аргументу, то в первом случае функция смещается на такую же величину, а во втором — «изменения нечувствительны». Скажем, если $x = 0{,}01$ (а это еще можно изобразить на миллиметровой бумаге), то $f_2(x) = 0{,}0001$, а это уже совершенно «нечувствительно».

Вот так обстоит дело с теоремой Ферма. Некоторые сведения исторического характера мы отложим до четырнадцатого рассказа.

Переходим к теореме Вейерштрасса о существовании экстремума непрерывной функции на конечном отрезке. Вследствие того, что любой отрезок можно перевести аффинным преобразованием в единичный отрезок $[0, 1]$, далее всюду будем рассматривать именно его.

Предварительно докажем следующую лемму о монотонной последовательности чисел.

Лемма. *Монотонная последовательность чисел из единичного отрезка имеет предел в этом отрезке.*

Иначе говоря, если задана бесконечная последовательность чисел $\{x_1, \ldots, x_n, \ldots\}$, обладающих теми свойствами, что, во-первых, они *из единичного отрезка*, т. е. $0 \leqslant x_n \leqslant 1$, $n = 1, 2, \ldots$, а во-вторых, что они *монотонны*, — для определенности, пусть они монотонно возрастают, т. е. $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n \leqslant \ldots$, то существует число x_0 из единичного отрезка ($0 \leqslant x_0 \leqslant 1$) такое, что $\lim\limits_{n\to\infty} x_n = x_0$.

Прежде чем доказывать лемму, необходимо сказать, что такое *число из единичного отрезка*. Числа из единичного отрезка представляются бесконечными десятичными дробями $0{,}n_1n_2n_3\ldots$, где n_i — одна из десяти цифр 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

Доказательство. Возьмем первые числа после запятой десятичного представления чисел последовательности $\{x_1, \ldots$ $\ldots, x_n, \ldots\}$. Получим (вследствие монотонности нашей последовательности) возрастающую последовательность целых чисел, каждое из которых не меньше нуля и не больше девяти. Какое-то из этих целых чисел, обозначим его n_1, *должно повториться бесконечное число раз*. Возьмем тот наименьший номер N_1, когда это целое число встретится у нас впервые. *Тогда чисел, бóльших n_1, в нашей последовательности уже не встретится*, так как если бы нашлось какое-то бóльшее число, то из-за монотонности этой последовательности число n_1 далее уже не встретилось бы.

Далее возьмем вторые числа после запятой последовательности $\{x_{N_1}, x_{N_1+1}, \ldots\}$. Снова получим возрастающую последовательность целых чисел, каждое из которых не меньше нуля и не больше девяти, и снова возьмем число n_2, которое встретится бесконечное число раз, и номер N_2, когда оно появится впервые. Далее поступаем аналогично. В итоге приходим к десятичной дроби $0{,}n_1n_2\ldots$ Она представляет собой некоторое число x_0 из единичного отрезка. Кроме того, у нас возникла последовательность номеров $N_1 \leqslant N_2 \leqslant \ldots \leqslant N_s \ldots$, где числа n_s появляются впервые. Из построения ясно, что для любого $n = 1, 2, \ldots$ выполнено неравенство

$$x_n \leqslant x_0 \iff x_n - x_0 \leqslant 0.$$

С другой стороны, если $n \geqslant N_s$, то

$$x_0 - x_n \leqslant 10^{-s}.$$

Написанные соотношения означают, что $\lim\limits_{n\to\infty} x_n = x_0$. Лемма доказана.

Т е о р е м а В е й е р ш т р а с с а. *Непрерывная функция на конечном отрезке принимает свое максимальное и минимальное значение.*

Напомним, что функция $y = f(x)$, определенная на отрезке $[a, b]$, содержащем точку x_0 ($a \leqslant x_0 \leqslant b$), называется *непрерывной в точке* x_0, если для любого $\varepsilon > 0$ можно указать $\delta > 0$ такое, что из неравенств $|x - x_0| < \delta$, $a \leqslant x \leqslant b$, последует неравенство $|f(x) - f(x_0)| < \varepsilon$. Из этого определения сразу следует, что если f непрерывна в точке x_0 и $\{x_1, \ldots, x_n, \ldots\}$ — последовательность, сходящаяся к x_0 ($\lim\limits_{n\to\infty} x_n = x_0$), то последовательность $\{f(x_1), \ldots, f(x_n), \ldots\}$, сходится к $f(x_0)$,

$$\lim_{n\to\infty} f(x_n) = f(x_0).$$

Функция, непрерывная в каждой точке отрезка, называется *непрерывной на отрезке*. Говорят, что в точке x_0 функция $y = f(x)$, заданная на $[a, b]$, принимает на нем свое максимальное (минимальное) значение, если $f(x_0) \geqslant f(x)$ ($f(x_0) \leqslant f(x)$) для любого x из отрезка $[a, b]$.

Теперь можно перейти к доказательству теоремы Вейерштрасса. Будем доказывать ее для максимума.

Д о к а з а т е л ь с т в о. Пусть функция $y = f(x)$ определена и непрерывна на единичном отрезке $[0, 1]$. Возьмем два отрезка: $\Delta_1 = [a_1, b_1]$ и $\Delta_2 = [a_2, b_2]$, содержащиеся в $[0, 1]$. Мы скажем, что отрезок Δ_1 *лучше* Δ_2, если найдется такая точка $\bar{x}$ из Δ_1, что $f(\bar{x}) > f(x)$ для любого x из Δ_2.

Разобьем отрезок $\Delta^0 = [0,1]$ на два равных отрезка: $\Delta_1^1 = [0,1/2]$ и $\Delta_2^1 = [1/2,1]$.

Выберем из отрезков Δ_1^1 и Δ_2^1 лучший; если ни один из них не лучше другого, возьмем любой из этих отрезков. Левую точку выбранного отрезка Δ^1 обозначим x_1.

В силу нашего выбора для всякой точки x, не лежащей в Δ^1, найдется точка $\bar{x}$ из Δ^1, в которой $f(\bar{x}) \geqslant f(x)$.

Действительно, если Δ^1 — лучший, то все очевидно, а если он не лучший и точка $\bar{x}$ не найдется, то это означает, что другой отрезок лучший, что противоречит нашему выбору.

Далее разобьем отрезок Δ^1 на два равных отрезка Δ_1^2 и Δ_2^2 и снова выберем либо лучший, либо любой, если лучшего нет. Левую точку выбранного отрезка Δ^2 обозначим x_2. Снова в силу нашего выбора для всякой точки x, не лежащей в Δ^2, найдется точка $\bar{x}$ из Δ^2, в которой $f(\bar{x}) \geqslant f(x)$ (продумайте).

Далее будем поступать аналогично. В итоге получаем монотонную последовательность $\{x_1, \ldots, x_n, \ldots\}$ из $[0,1]$. По лемме она имеет предел, который обозначим x_0. Докажем, что $f(x_0) \geqslant f(x)$ для всякого x из $[0,1]$. Действительно, допустим, что существует точка $\widetilde{x}$, в которой $f(\widetilde{x}) > f(x_0)$. Выберем δ столь малым, чтобы $|x_0 - \widetilde{x}| > \delta$ и из неравенств $|x - x_0| < \delta$, $0 \leqslant x \leqslant 1$ следовало бы, что $f(x) < f(\widetilde{x})$. Длины отрезков Δ^n равны 2^{-n}, а их левые концы x_n стремятся к x_0. Значит, в какой-то момент весь отрезок Δ^n окажется внутри интервала $(x_0 - \delta, x_0 + \delta)$. Но тогда, с одной стороны, в нем найдется $\bar{x}$ такое, что $f(\bar{x}) \geqslant f(\widetilde{x})$, а с другой стороны (так как $|\bar{x} - x_0| < \delta$), $f(\bar{x}) < f(\widetilde{x})$. Пришли к противоречию. Теорема доказана.

Следствие из теоремы (сформулированное в начале рассказа) доказывается совсем просто.

Найдем такое число A, чтобы для $|x| \geqslant A$ выполнялось неравенство $f_0(x) \geqslant f_0(0)$. По теореме Вейерштрасса найдется точка x_0 из отрезка $[-A, A]$ такая, что $f(x_0) \leqslant f(x)$ для любого x из отрезка $[-A, A]$. Ясно, что $f(x_0) \leqslant f(0)$. Но если $|x| > A$, то $f(x) \geqslant f(0) \geqslant f(x_0)$. Значит, $f(x) \geqslant f(x_0)$ для любого x, что и требовалось получить. Два других случая доказываются так же просто.

Таким образом, мы привели доказательство всех фактов, о которых речь шла в п. 1.

Рассказ двенадцатый

ЭКСТРЕМУМЫ ФУНКЦИЙ МНОГИХ ПЕРЕМЕННЫХ. ПРИНЦИП ЛАГРАНЖА

Можно высказать следующий общий принцип. Если ищется максимум или минимум некоторой функции многих переменных при условии, что между этими переменными имеется связь, задаваемая одним или несколькими уравнениями, нужно прибавить к функции, экстремум которой мы ищем, функции, задающие уравнения связи, умноженные на неопределенные множители, и искать затем максимум или минимум построенной суммы, как если бы переменные были независимы. Полученные уравнения, присоединенные к уравнениям связи, послужат для определения всех неизвестных.

Ж. Лагранж

1. Здесь будет рассказано о том, как решать задачи на экстремум функций многих переменных. Суть дела выражена Лагранжем в словах, поставленных нами в качестве эпиграфа. Для освоения первой части этого рассказа необходимо владеть понятиями «непрерывная функция многих переменных» (рассказ десятый) и материалом первой части предыдущего рассказа.

Пусть $f_0, f_1, \ldots, f_m$ — функции n переменных $x = (x_1, \ldots, x_n)$. Рассмотрим задачу с ограничениями типа равенств и неравенств

$$(\text{з}) \qquad f_0(x) \to \min\ (\max), \quad f_i(x) = 0,\ i = 1, \ldots, m$$

$$f_i(x) \leqslant 0, \quad i = m' + 1, \ldots, m.$$

Отметим, впрочем, что в основном у нас будут встречаться задачи с одними лишь равенствами.

П р и м е р ы.

$$(\text{з}_1) \quad f_0(x) = (x_1 - a_1)^2 + (x_2 - a_2)^2 + (x_1 - b_1)^2 + (x_2 - b_2)^2 + \\ + (x_1 - c_1)^2 + (x_2 - c_2)^2 \to \min, \quad x = (x_1, x_2).$$

$$(\text{з}_2) \quad f_0(x) = x_1 \ldots x_n \to \max, \quad f_1(x) = x_1^2 + \ldots + x_n^2 - 1 = 0.$$

Напомним, что (з_2) — это при $n = 2$ формализация планиметрической, а при $n = 3$ — классической задачи Кеплера о параллелепипеде максимального объема, вписанном в шар.

(з$_1$) формализует такую задачу: *найти на плоскости точку, сумма квадратов расстояний от которой до трех данных точек была бы минимальной* (сравните ее с задачей Штейнера).

Разумеется, не каждая задача типа (з) имеет решение. Однако, как и для функций одного переменного, можно сформулировать теорему (также доказанную Вейерштрассом), которая во многих случаях гарантирует существование решения.

Обозначим совокупность допустимых точек в задаче (з) через C. Иначе говоря, C состоит из тех точек x, для которых

$$f_i(x) = 0,\ i = 1, \ldots, m', \quad f_i(x) \leqslant 0,\ i = m'+1, \ldots, m.$$

Множество C называется ограниченным, если существует константа $A > 0$ такая, что $|x_i| \leqslant A$, $i = 1, \ldots, n$, для любой точки $x = (x_1, \ldots, x_n)$ из C. Например, множество $x_1^2 + \ldots + x_n^2 = 1$ ограничено, в частности, ограничена окружность $x_1^2 + x_2^2 = 1$, а множество $x_1 = x_2^2$ (парабола) не является ограниченным множеством.

Теорема Вейерштрасса. *Пусть в задаче* (з) *функции* $f_0, \ldots, f_m$ *непрерывны, а совокупность* C *допустимых точек в* (з) *ограничена. Тогда решение обеих задач*

(з$_{\min}$)
$$f_0(x) \to \min, \quad f_i(x) = 0,\ i = 1, \ldots, m', \quad f_i(x) \leqslant 0,\ i = m'+1, \ldots, m,$$

и

(з$_{\max}$)
$$f_0(x) \to \max, \quad f_i(x) = 0,\ i = 1, \ldots, m', \quad f_i(x) \leqslant 0,\ i = m'+1, \ldots, m,$$

существует.

Укажем, как и в предыдущем рассказе, на такое следствие.

Следствие. *Пусть функция* $f_0(x)$, $x = (x_1, \ldots, x_n)$, *непрерывна для любого* x. *Тогда, если*

$$\lim f_0(x) = \infty \quad \textit{при} \quad x_1^2 + \ldots + x_n^2 \to \infty,$$

то решение задачи без ограничений

$$f_0(x) \to \min$$

существует.

Повторим определение локального экстремума в задаче (з).

Определение 1. Говорят, что точка $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ доставляет *локальный минимум* (*максимум*) в задаче (з), если можно указать такое $\varepsilon > 0$, что для всех допустимых точек $x = (x_1, \ldots \ldots, x_n)$, удовлетворяющих неравенству

$$|x_i - \widehat{x}_i| < \varepsilon, \quad i = 1, \ldots, n,$$

выполнено неравенство $f_0(x) \geqslant f_0(\widehat{x})$ $(f_0(x) \leqslant f_0(\widehat{x}))$.

Если (з) — задача без ограничений и $\widehat{x}$ доставляет ей локальный экстремум, мы говорим еще, что $\widehat{x}$ доставляет локальный экстремум функции f_0.

Для того чтобы можно было сформулировать основное правило решения задач типа (з), необходимо ввести еще одно понятие.

Пусть функция $y = f(x)$ определена для всех $x = (x_1, \ldots, x_n)$, для которых выполнены неравенства $a_j \leqslant x_j \leqslant b_j$, $j = 1, \ldots, n$ (совокупность всех таких точек назовем параллелепипедом $\Pi(a_1, b_1; \ldots; a_n, b_n)$), и пусть для точки $x_0 = (x_{01}, \ldots, x_{0n})$ выполнены строгие неравенства

$$a_j < x_{0j} < b_j, \quad j = 1, \ldots, n.$$

Рассмотрим следующую функцию одной переменной:

$$g_j(x) = f(x_{01}, \ldots, x_{0,j-1}, x_{0,j} + x, x_{0,j+1}, \ldots, x_{0n}).$$

Что мы сделали? Мы закрепили все координаты у x_0, кроме j-й, а к j-й координате прибавили число x. Предположим теперь, что функция g_j дифференцируема в нуле.

Определение 2. Производная в нуле функции g_j называется j-й *частной производной функции f в точке* x_0 и обозначается $\partial f(x_0)/\partial x_j$.

Теперь сформулируем две теоремы, которые позволят нам затем описать правило решения задач типа (з).

Сначала рассмотрим задачу (з) без ограничений.

Теорема Ферма. *Пусть в точке $\widehat{x}$ существуют все частные производные функции f_0. Если точка $\widehat{x}$ доставляет локальный экстремум* (*минимум или максимум*) *этой функции, то*

$$\frac{\partial f_0(\widehat{x})}{\partial x_j} = 0, \quad j = 1, \ldots, n. \tag{1}$$

Точки, для которых все частные производные равны нулю, называются *стационарными*. Соотношения (1) являются, разумеется (как и в случае $n = 1$), *необходимыми условиями*.

Приведем пример приложения этой теоремы. Решим задачу ($з_1$). Решение задачи существует в силу следствия теоремы Вейерштрасса. Обозначим его $(\widehat{x}, \widehat{y})$. Имеем

$$0 = \frac{\partial f_0(\widehat{x}, \widehat{y})}{\partial x_1} = 2[(\widehat{x} - a_1) + (\widehat{x} - b_1) + (\widehat{x} - c_1)] \Longrightarrow \widehat{x} = (a_1 + b_1 + c_1)/3,$$

$$0 = \frac{\partial f_0(\widehat{x}, \widehat{y})}{\partial x_2} = 2[(\widehat{y} - a_2) + (\widehat{y} - b_2) + (\widehat{y} - c_2)] \Longrightarrow \widehat{y} = (a_2 + b_2 + c_2)/3.$$

О т в е т: $(\widehat{x}, \widehat{y})$ есть центр масс треугольника (a_1, a_2), (b_1, b_2), (c_1, c_2).

Теперь рассмотрим общую задачу (з), в которой, правда, *неравенства отсутствуют*. Составим такую сумму:

$$\mathscr{L} = \lambda_0 f_0(x) + \lambda_1 f_1(x) + \ldots + \lambda_m f_m(x),$$

где $x = (x_1, \ldots, x_n)$, $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)$.

Эту функцию называют *функцией Лагранжа*. Числа $\lambda_0, \lambda_1, \ldots$ $\ldots, \lambda_m$ называются *множителями Лагранжа*.

Общий принцип Лагранжа, описание которого вынесено нами в эпиграф, можно коротко выразить так: *для решения задачи* (з) (с одними равенствами) *надо составить функцию Лагранжа, а затем поступить с ней так, как если бы переменные* $(x_1, \ldots, x_n)$ *были независимы* (т. е. применить теорему Ферма). Полученные уравнения, а именно уравнения

$$\frac{\partial \mathscr{L}(x, \lambda)}{\partial x_j} = 0, \quad j = 1, \ldots, n, \tag{2}$$

дополненные уравнениями связи

$$f_i(x) = 0, \quad i = 1, \ldots, m, \tag{3}$$

надо решить относительно $x_1, \ldots, x_n$, $\lambda_0, \ldots, \lambda_m$ и среди этих решений выбрать нужное.

Это правило основывается на следующей теореме.

Т е о р е м а (п р а в и л о м н о ж и т е л е й Л а г р а н ж а). *Пусть функции $f_0, \ldots, f_m$, определены в некотором параллелепипеде* $\Pi(a_1, b_1; a_2, b_2; \ldots; a_n, b_n)$, *содержащем внутри себя точку* $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ $(a_i < \widehat{x}_i < b_i,\ i = 1, \ldots, n)$. *Пусть далее все функции f_i, $i = 0, 1, \ldots, m$, и все частные производные $\dfrac{\partial f_i}{\partial x_j}$, $i = 0, 1, \ldots$ $\ldots, m$; $j = 1, \ldots, n$, непрерывны в этом параллелепипеде. Тогда, если допустимая точка $\widehat{x}$ доставляет локальный экстремум* (*минимум или максимум*) *в задаче* (з) (*где неравенства отсутствуют*: $m' = m$), *то найдутся числа* $\lambda_0, \ldots, \lambda_m$, *не равные одновременно нулю и такие, что*

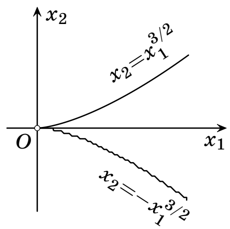
$$\frac{\partial \mathscr{L}(\widehat{x}, \lambda)}{\partial x_j} = 0, \quad j = 1, \ldots, n$$

$$(\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n), \quad \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m)).$$

Сделаем теперь несколько замечаний. Отметим, во-первых, что в (2) и (3) имеется $n + m$ уравнений с $n + m + 1$ неизвестным. Но надо иметь в виду, *что множители Лагранжа можно умножать на любую константу, отличную от нуля*.

Таким образом, всегда можно домножить так, чтобы один из множителей Лагранжа равнялся единице. Тогда в (2)—(3) *число уравнений фактически равно числу неизвестных*.

Далее. Уравнения (2) наиболее содержательны, если $\lambda_0 \neq 0$. Ведь если $\lambda_0 = 0$, то уравнения (2) отражают лишь вырожденность ограничений и не связаны с функцией, экстремум которой ищется. Но вообще-то заранее считать, что λ_0 отлично от нуля (как это делает Лагранж — см. эпиграф), нельзя. Приведем пример, показывающий, что правило множителей Лагранжа при дополнительном допущении $\lambda_0 \neq 0$ может стать неверным.



П р и м е р. Рассмотрим задачу (см. рис. 48):

$$x_1 \to \min, \quad x_2^2 - x_1^3 = 0.$$

Из рисунка видно, что единственным решением задачи является точка $\widehat{x} = (0, 0)$. Попробуем теперь составить функцию Лагранжа с $\lambda_0 = 1$ и применить алгоритм Лагранжа:

$$\mathscr{L} = x_1 + \lambda(x_2^2 - x_1^3),$$

откуда

$$\frac{\partial \mathscr{L}}{\partial x_1} = 0 \implies -3\lambda x_1^2 + 1 = 0, \quad \frac{\partial \mathscr{L}}{\partial x_2} = 0 \implies 2\lambda x_2 = 0$$

Уравнение связи имеет вид $x_2^2 - x_1^3 = 0$. Написанные уравнения, как легко видеть, несовместны.

Правило множителей Лагранжа позволяет дать следующий рецепт поиска решений задачи (з) с равенствами. Разобьём его на те же 4 этапа.

Этап 1 — формализация задачи. Требуется привести (разумеется, если это возможно) задачу к виду (з) с $m' = m$.

Этап 2 — применение принципа Лагранжа, т. е. составление системы уравнений

$$\frac{\partial \mathscr{L}}{\partial x_j} = 0, \quad j = 1, \ldots, n,$$
$$f_i(x) = 0, \quad i = 1, \ldots, m.$$

Этап 3 состоит в нахождении всех стационарных точек. При этом бывает полезно сначала выяснить вопрос о том, может ли λ_0 равняться нулю или нет.

Этап 4 состоит в переборе всех стационарных точек и выборе тех, где функция f_0 имеет минимальное (максимальное) значение. Этот рецепт решения мы коротко называем *принципом Лагранжа*.

Из сформулированных теорем следует, что *если* (в задаче без неравенств) *совокупность допустимых точек ограничена, все функции* $f_0, \ldots, f_m$ *непрерывны, и непрерывны все частные производные* $\frac{\partial f_i}{\partial x_j}$, $i = 0, \ldots, m$, $j = 1, \ldots, n$, *то описанное правило приведет к решению задачи.*

Искусству дифференцировать функции многих переменных заново обучать не надо: ведь все сводится к дифференцированию функций одной переменной. Теперь уже можно было бы перейти к решению задач. Но мы потратим еще некоторое время на обсуждения.

2. Здесь так же, как и в предыдущем рассказе, уделим место некоторым объяснениям.

Теорему Ферма в многомерном случае нет нужды специально объяснять, ибо она тривиально выводится из одномерного случая. Действительно, пусть функция $f_0(x_1, \ldots, x_n)$ имеет локальный экстремум в точке $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$. То есть функция $g_j(x)$ (по типу той, которая была определена в п. 1):

$$g_j(x) = f_0(\widehat{x}_1, \ldots, \widehat{x}_{j-1}, \widehat{x}_j + x, \widehat{x}_{j+1}, \ldots, \widehat{x}_n)$$

должна иметь минимум в нуле. Но тогда по одномерной теореме Ферма, которая подробно освещалась в предыдущем рассказе, получаем

$$g_j'(0) = 0.$$

Но по самому определению j-й частной производной как раз и получается, что

$$g_j'(0) = \frac{\partial f_0(\widehat{x})}{\partial x_j}.$$

Сопоставляя эти два равенства и произвольность j, приходим к теореме Ферма:

$$\frac{\partial f_0(\widehat{x})}{\partial x_j} = \frac{\partial f_0(\widehat{x}_1, \ldots, \widehat{x}_n)}{\partial x_j} = 0, \quad j = 1, \ldots, n.$$

Теперь нам предстоит разобраться в правиле множителей Лагранжа. Начнем наши объяснения с простейшей ситуации, когда, во-первых, $n = 2$, $m = 1$, а во-вторых, когда ограничение задается *линейным соотношением*. Иначе говоря, будем рассматривать следующую задачу:

$$f_0(x_1, x_2) \to \max, \quad f_1(x_1, x_2) = a_1 x_1 + a_2 x_2 - b = 0.$$

График $y = f_0(x_1, x_2)$ можно мыслить себе как ландшафт горной местности. Соотношение $f_1(x_1, x_2) = a_1 x_1 + a_2 x_2 - b = 0$ задает на плоскости прямую. Можно представить себе, что в горной местности прокладывают линию электропередачи по траектории, которая на карте имеет вид прямой. Спрашивается, где будет самая высокая точка трассы?

Вспомним, как горы изображают на картах. Там наносят линии одинаковой высоты, линии уровня. Задумаемся над тем, как должны быть расположены друг относительно друга трасса электропередачи и линия уровня нашей горы в точке наивысшего подъема.

Ясно, что *трасса в этой точке не может пересекать линию уровня*. Если трасса пересекает ее, то она проходит из области меньших значений высоты в область ее больших значений, и значит, в точке пересечения максимума высоты быть не может. Вывод: *в точке максимума высоты трасса должна касаться линии уровня*.

Теперь посмотрим на функцию $f_1(x_1, x_2) = a_1 x_1 + a_2 x_2 - b$. Ее частные производные

$$\frac{\partial f_1}{\partial x_1} = a_1, \quad \frac{\partial f_1}{\partial x_2} = a_2.$$

Вектор (a_1, a_2) перпендикулярен трассе и вообще любой линии уровня функции $f_1(x_1, x_2)$. Это ясно из геометрического смысла уравнения $a_1 x_1 + a_2 x_2 = b$. Однако оказывается, что этот факт верен всегда. А именно, если $f_1(x_1, x_2)$ — любая непрерывная функция, у которой ее частные производные также непрерывны, то вектор $\left(\frac{\partial f(\bar{x})}{\partial x_1}, \frac{\partial f(\bar{x})}{\partial x_2}\right)$ *перпендикулярен касательной к линии уровня* $f(x) = f(\bar{x})$. *Но ведь трасса тоже касательна* *к линии уровня в точке максимума высоты.* Из всего сказанного и вытекает, что оба вектора $\left(\dfrac{\partial f_0(\widehat{x})}{\partial x_1}, \dfrac{\partial f_0(\widehat{x})}{\partial x_2}\right)$ и (a_1, a_2) перпендикулярны трассе и, следовательно, пропорциональны, т. е.

$$\frac{\partial f_0(\widehat{x})}{\partial x_1} + \lambda a_1 = 0, \quad \frac{\partial f_0(\widehat{x})}{\partial x_2} + \lambda a_2 = 0$$

или

$$\frac{\partial \mathscr{L}(\widehat{x}, 1, \lambda)}{\partial x_1} = 0, \quad \frac{\partial \mathscr{L}(\widehat{x}, 1, \lambda)}{\partial x_2} = 0,$$

где

$$\mathscr{L}(x, \lambda_0, \lambda) = \lambda_0 f_0(x) + \lambda f_1(x).$$

В этой частной ситуации мы пришли к правилу множителей Лагранжа.

Пусть теперь функция f_1 — не обязательно аффинная, но у нее имеются непрерывные частные производные. Снова рассмотрим семейство линий уровня

$$f_0(x_1, x_2) = c_0, \quad f_1(x_1, x_2) = c_1.$$

Допустим, что в некоторой части плоскости через каждую точку проходит одна и лишь одна кривая каждого из семейств. Предположим, что в $\widehat{x} = (\widehat{x}_1, \widehat{x}_2)$ задача имеет решение. Тогда все точки кривой l_0, задаваемой уравнением

$$f_0(x_1, x_2) = \widehat{c}_0 =: f_0(\widehat{x}_1, \widehat{x}_2),$$

обязаны лежать «по одну сторону» от кривой l_1, задаваемой уравнением

$$f_1(x_1, x_2) = \widehat{c}_1 =: f_1(\widehat{x}_1, \widehat{x}_2),$$

т. е. эти кривые должны не пересекаться, а касаться. Мы объяснили, что если в точке $\widehat{x}$ на кривой l_1 функция f_0 имеет экстремальное значение, то кривая l_0 касается кривой l_1.

Еще раз напомним о том, что вектор $(\partial f_0(\widehat{x})/\partial x_1, \partial f_0(\widehat{x})/\partial x_2)$ перпендикулярен кривой l_0, а вектор $(\partial f_1(\widehat{x})/\partial x_1, \partial f_1(\widehat{x})/\partial x_2)$ перпендикулярен кривой l_1. Если эти кривые касаются, то данные векторы должны быть пропорциональны.

Пропорциональность этих векторов и утверждается в правиле множителей Лагранжа (с $\lambda_0 = 1$).

Доказать принцип Лагранжа (т. е. правило множителей Лагранжа) в этой книге мы не сможем. Однако некоторые пояснения, касающиеся доказательства его, будут даны в четырнадцатом рассказе.

В начале рассказа была поставлена общая задача, когда имеются и равенства, и неравенства. А дальше речь шла уже об одних лишь равенствах. Как видоизменяется принцип Лагранжа в общем случае? Вернемся к задаче (з), которая была поставлена вначале, и допустим, что все функции $f_0(x), \ldots, f_m(x)$ удовлетворяют условиям сформулированной выше теоремы о правиле множителей Лагранжа. Тогда, если допустимая точка $\widehat{x}$ доставляет локальный минимум в задаче (з), то найдутся числа $\lambda_0, \lambda_1, \ldots, \lambda_m$, не равные одновременно нулю и такие, что $\dfrac{\partial \mathscr{L}(\widehat{x}, \lambda)}{\partial x_j} = 0$, $j = 1, \ldots, n$ $(\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$, $\lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m))$, кроме того, удовлетворяются условия неотрицательности множителей Лагранжа при функционале и неравенствах $\lambda_0 \geqslant 0$, $\lambda_{m'+1} \geqslant 0$, $\ldots$, $\lambda_m \geqslant 0$ и выполняются следующие условия (так называемые *условия дополняющей нежесткости*):

$$\lambda_j f_j(\widehat{x}) = 0, \quad j = m' + 1, \ldots, m.$$

Условия дополняющей нежесткости означают, что множитель Лагранжа λ_j, может быть отличен от нуля только на «активном» ограничении, когда в точке экстремума ограничение типа неравенства на самом деле является равенством: $f_j(\widehat{x}) = 0$, $j = m' + 1, \ldots, m$.

Что же изменилось? Во-первых, здесь «экстремум» заменился на «минимум». При неравенствах небезразлично, какой тип экстремума рассматривается. В случае задачи на максимум или при наличии ограничений вида $f_j \geqslant 0$ для применения только что сформулированного результата необходимо сначала преобразовать задачу к виду (з), заменив, если это нужно, некоторые из f_j, $j = 0, m' + 1, \ldots, m$, на $-f_j$. Во-вторых, добавилось условие неотрицательности, в-третьих — условие дополняющей нежесткости.

При решении задач в этой книге нам не придется применять только что сформулированную теорему. Но роль ее и ей подобных теорем очень значительна. В первом рассказе мы както сказали, что старые методы оказались недостаточными для решения многих экономических проблем. Теперь об этом можно сказать конкретнее.

Вспомним транспортную задачу, о которой разговор заходил уже дважды. В ее формализации участвуют неравенства, без них формализировать ее затруднительно. Но раньше задач с неравенствами не рассматривали. То дополнение к правилу множителей Лагранжа, которое только что было сделано, появилось лишь шестьдесят (а не двести пятьдесят) лет тому назад. И еще одно. Выяснилось, что в экономических задачах очень часто минимизируемые функции и ограничения выпуклы или даже линейны.

Пришлось специально заняться изучением выпуклых функций и выпуклых экстремальных задач. И мы уделим этому некоторое внимание.

3. Выпуклые функции и выпуклые задачи. О выпуклых функциях одной переменной мы мельком уже сказали. Выпуклые функции нескольких переменных определяются совершенно аналогично. Функция $f(x) = f(x_1, \ldots, x_n)$ называется *выпуклой*, если для любых точек x и x' и любого α, $0 \leqslant \alpha \leqslant 1$, имеет место неравенство Иенсена

$$f(\alpha x + (1-\alpha)x') \leqslant \alpha f(x) + (1-\alpha)f(x').$$

Примерами выпуклых функций являются прежде всего линейные и аффинные функции. Отметим также «*функцию расстояния*» от точки до нуля

$$f(x_1, \ldots, x_n) = \sqrt{x_1^2 + \ldots + x_n^2}.$$

Функцию $y = f(x)$ называют *строго выпуклой*, если в неравенстве Иенсена при $x \neq x'$ и $0 < \alpha < 1$ имеет место строгое неравенство

$$f(\alpha x + (1-\alpha)x') < \alpha f(x) + (1-\alpha)f(x').$$

Функции $y = x^2$, $y = \sqrt{h^2 + x^2}$ и $y = x_1^2 + x_2^2$ строго выпуклы, а функция $y = |x|$ или функция расстояния — не строго выпуклы. Нетрудно понять, что если функция $y = f(x)$ строго выпукла и достигает в точке $\widehat{x}$ своего минимума, то этот минимум единствен.

Уже было сказано, что не всякая функция является дифференцируемой, и дважды обсуждался пример: $y = |x|$. Эта функция выпукла и не дифференцируема в нуле. Функция расстояния также дифференцируема всюду, кроме нуля, где она не является дифференцируемой. Однако если выпуклая функция дифференцируема и в некоторой точке ее производная равна нулю, то в этой точке функция достигает своего абсолютного экстремума. И даже более того: график любой выпуклой функции лежит «по одну сторону» любой своей касательной плоскости.

Отмеченное обстоятельство очень важно: для выпуклых дифференцируемых функций теорема Ферма является достаточным условием экстремума! Вот одна из причин того, что теория выпуклых экстремальных задач имеет такой завершенный характер.

Рассказ тринадцатый

СНОВА ПОРЕШАЕМ!

Именно здесь мы намереваемся выполнить обещание, которое (возможно, неосторожно) взяли на себя ранее. Наша цель — перерешать все задачи из первой части, и при этом «единообразно, стандартно, если хотите — рутинно», применяя один и тот же (и притом простой) прием, а именно принцип Лагранжа или — в частном случае — теорему Ферма. (В трех случаях нам это сделать все-таки не удастся: мы не сможем справиться с классической изопериметрической задачей, с брахистохроной и задачей Ньютона. Чтобы стандартно решать их, надо будет еще потрудиться.)

При нашем стандартном исследовании выделяются четыре этапа: 1) формализация; 2) применение принципа Лагранжа или теоремы Ферма; 3) решение соответствующих уравнений и нахождение критических или стационарных точек; 4) отбор нужных точек и обсуждение ответа.

Начнем с задач, сводящихся к нахождению экстремумов функций одного переменного.

1. Задача Евклида о параллелограмме наибольшей площади, вписанном в треугольник (рассказ четвертый).

1°. Формализация. Снова обратимся к рис. 12. Пусть, как и раньше, H — высота треугольника ABC, b — длина AC. Обозначим длину AF' через x. Тогда $0 \leqslant x \leqslant b$. Через $h = h(x)$ обозначим высоту треугольника $BD'E'$. Из подобия треугольников $BD'E'$ и ABC ($D'E' || AC$) получаем $h(x)/H = x/b$. Площадь параллелограмма $AD'E'F'$ равна $(H - h(x))x = H(b-x)x/b$. В итоге приходим к следующей формализации:

$$(\text{з}_1) \qquad f_0(x) = \frac{Hx(b-x)}{b} \to \max, \quad 0 \leqslant x \leqslant b.$$

2°. Необходимое условие: $f_0'(x) = 0$.

3°. Нахождение критических точек. Стационарные точки: $f_0'(x) = (H(bx - x^2)/b)' = (b - 2x)H/b$, т. е. $f_0'(x) = 0$ лишь в точке $b/2$. Критические точки: 0, $b/2$, b.

4°. Функция f_0 непрерывна, дифференцируема всюду и рассматривается на конечном отрезке. Значит, решение находится среди критических точек. Перебор критических точек: $f_0(0) = f_0(b) = 0$, $f_0(b/2) > 0$.

Таким образом, решение ($з_1$) есть $b/2$. Ответ: *искомый параллелограмм $ADEF$ характеризуется тем, что точка F является серединой отрезка* $[AC]$. Именно этот факт и был установлен Евклидом.

2. Задача Архимеда о шаровом сегменте наибольшего объема среди изопифанных сегментов (рассказ четвертый).

1°. Формализация. Пусть R — радиус шара, а h — высота шарового сегмента. Объем шарового сегмента, как известно, равен $\pi h^2(R - h/3)$, а площадь его боковой поверхности — $2\pi Rh$. Площадь боковой поверхности задана: $2\pi Rh = a$, откуда $R = a/2\pi h$. Подставляя это выражение вместо R в формулу для объема и учитывая, что $h \leqslant 2R = a/\pi h$, получаем следующую формализацию:

$$(з_2) \qquad f_0(h) = \frac{ha}{2} - \frac{\pi h^3}{3} \to \max, \quad 0 \leqslant h \leqslant \sqrt{\frac{a}{\pi}}.$$

2°. Необходимое условие: $f_0'(h) = 0$.

3°. Нахождение критических точек. Стационарные точки: $f_0'(h) = (ha/2 - \pi h^3/3)' = a/2 - \pi h^2$, т. е. $f_0'(h) = 0$ лишь в случае, когда $h = \sqrt{a/2\pi}$. Критические точки: $0, \sqrt{a/2\pi}, \sqrt{a/\pi}$.

4°. Функция f_0 непрерывна, дифференцируема всюду и рассматривается на конечном отрезке. Значит, решение находится среди критических точек.

Перебор критических точек: $f_0(0) = 0$, $f_0(\sqrt{a/2\pi}) = \sqrt{2}a^{3/2}/6\sqrt{\pi}$, $f_0(\sqrt{a/\pi}) = a^{3/2}/6\sqrt{\pi}$. Максимальное значение доставляет точка $\sqrt{a/2\pi}$. Это и есть решение. Вспомнив, что $a = 2\pi Rh$, получаем $h = R$. Ответ: *искомый шаровой сегмент является полушаром — высота равна радиусу*. Именно этот результат и был доказан Архимедом.

3. Задача о наименьшей площади (рассказ четвертый).

1°. Формализация. Снова обратимся к рис. 17. Проведем через точку M прямую, параллельную AB, и точку пересечения этой прямой с AC обозначим через N. Пусть E' — некоторая точка на луче NC, $a = |AN|$, $x = |NE'|$, D' — точка пересечения луча AB и прямой $E'M$. Треугольники $ME'N$ и $AD'E'$ подобны, ибо $MN \parallel AB$. Значит, их площади относятся как квадраты длин отрезков $[NE']$ и $[AE']$. Но площадь треугольника $ME'N$ равна $xh/2$, где h — высота, опущенная из M на AC. Отсюда получаем, что площадь треугольника $AD'E'$ равна $h(x+a)^2/2x$, и мы приходим к такой формализации:

$$(з_3) \qquad f_0(x) = \frac{(a+x)^2}{x} \to \min, \quad x > 0.$$

2°. Необходимое условие: $f_0'(x) = 0$.

3°. Нахождение стационарных точек:

$$f_0'(x) = \left(\frac{(a+x)^2}{x}\right)' = \left(\frac{a^2}{x} + 2a + x\right)' = -\frac{a^2}{x^2} + 1,$$

т. е. $f_0'(x) = 0$ лишь в случае $x = a$.

4°. Функция f_0 удовлетворяет требованиям следствия из теоремы Вейерштрасса из рассказа одиннадцатого. Значит, *решение задачи* (з$_3$) *существует*. В силу дифференцируемости f_0 для $x > 0$ и теоремы Ферма *решение должно быть стационарной точкой*. Но стационарная точка *единственна*. Значит, она и является решением. Ответ: *искомая точка E удалена от A на расстояние* $2a$. Отсюда сразу следует, что точка M делит отрезок DE пополам (из подобия треугольников ENM и EAD). Именно этот результат и был получен нами ранее геометрическим путем.

4. Задача Герона. Она уже много раз встречалась у нас — в первом, втором, десятом и одиннадцатом рассказах. Пришло время решить ее стандартно.

1°. Формализация уже была проведена в рассказе десятом:

$$(з_4) \qquad f_0(x) = \sqrt{a^2 + x^2} + \sqrt{b^2 + (d-x)^2} \to \min.$$

2°. Необходимое условие: $f_0'(x) = 0$.

3°. Нахождение стационарных точек. Применяя теорему о дифференцировании сложной функции, получим: $(\sqrt{a^2+x^2})' = x/\sqrt{a^2+x^2}$ (это было подробно проделано в одиннадцатом рассказе);

$$(\sqrt{b^2 + (d-x)^2})' = -\frac{d-x}{\sqrt{b^2+(d-x)^2}}.$$

Отсюда получаем

$$\frac{x}{\sqrt{a^2+x^2}} = \frac{d-x}{\sqrt{b^2+(d-x)^2}}. \tag{1}$$

Решим полученное уравнение. Возводя в квадрат, «переворачивая» и вычитая по единице, приходим к соотношению $(a/x)^2 = (b/(d-x))^2$, откуда (отбросив посторонний корень) получаем, что $x/a = (d-x)/b$ (это согласуется с тем, что говорилось в первом рассказе — см. рис. 1). Решение полученного уравнения обозначим через $\widehat{x}$.

4°. Функция $f_0(x)$ как сумма двух выпуклых функций выпукла, кроме того, она гладкая. Значит, как было сказано в конце предыдущего рассказа, $\widehat{x}$ — решение задачи, причем единственное в силу строгой выпуклости $f_0(x)$.

Взглянем еще раз на самый первый рисунок этой книги. Величина $\widehat{x}/\sqrt{a^2+\widehat{x}^2}$ равна $\sin\varphi_1$, величина $(d-\widehat{x})/\sqrt{b^2+(d-\widehat{x})^2}$ равна $\sin\varphi_2$. Из (1) следует, что $\sin\varphi_1 = \sin\varphi_2$, т. е. $\varphi_1 = \varphi_2$.

Ответ: *решение задачи Герона характеризуется тем, что угол падения равен углу отражения.* Этот факт и был доказан нами на первых же страницах этой книги.

В первом рассказе были сформулированы две задачи — №№ 1 и 2, примыкающие к задаче Герона. Но они тривиально к ней сводятся. Докажите, что в задаче 1 решение существует. Тогда, если оно не совпадает с вершиной угла, оно является решением задачи Герона для B, C и той стороны угла, где должна находиться точка A. Значит, у точки A угол падения равен углу отражения. Аналогичное утверждение верно и для точки B. Это приводит к нужному построению. Так же решается и задача 2, поэтому нет нужды решать их формально.

5. Задача Снеллиуса о законе преломления (рассказ третий). Пришло время и эту задачу решить стандартно — так, как это впервые сделал Лейбниц.

1°. Формализация. Направим ось Ox по прямой, разделяющей две среды, а ось Oy проведем перпендикулярно оси Ox через точку A (см. рис. 9). Пусть координаты точек A, B оказались такими: $A=(0,a)$, $B=(d,-b)$. Возьмем на оси точку D' с координатой $(x,0)$. Тогда время, затрачиваемое светом на преодоление пути $AD'B$, равно $\sqrt{a^2+x^2}/v_1 + \sqrt{b^2+(d-x)^2}/v_2$.

В итоге получается задача без ограничений:

$$(з_5) \qquad f_0(x) = \frac{\sqrt{a^2+x^2}}{v_1} + \frac{\sqrt{b^2+(d-x)^2}}{v_2} \to \min.$$

2°. Необходимое условие: $f_0'(x)=0$.

3°. Нахождение стационарных точек:

$$f_0'(x) = \left(\frac{\sqrt{a^2+x^2}}{v_1}\right)' + \left(\frac{\sqrt{b^2+(d-x)^2}}{v_2}\right)' = \\ = \frac{x}{v_1\sqrt{a^2+x^2}} - \frac{d-x}{v_2\sqrt{b^2+(d-x)^2}} = 0. \quad (1)$$

(По поводу дифференцирования см. предыдущую задачу.) Вследствие монотонности функций $x/v_1\sqrt{a^2+x^2}$ и $(d-x)/v_2\sqrt{b^2+(d-x)^2}$ получаем, что имеется единственное решение $\widehat{x}$ уравнения $f_0'(x)=0$.

4°. Функция $f_0(x)$ выпукла, значит, теорема Ферма является достаточным условием экстремума, т. е. $\widehat{x}$ — решение. Из строгой выпуклости $f_0(x)$ следует, что это решение единственно. Из рис. 9 и соотношения (1) получаем равенство

$$\frac{\sin\alpha_1}{v_1}=\frac{\sin\alpha_2}{v_2}, \tag{2}$$

выражающее закон Снеллиуса. Ответ: *решение в задаче Снеллиуса характеризуется тем, что отношение синусов угла падения и угла преломления равно отношению скоростей в первой и второй средах*. Этот факт был установлен нами в третьем рассказе.

6. Планиметрическая задача Кеплера о прямоугольнике максимальной площади, вписанном в круг. Об этой задаче говорилось в рассказах пятом и десятом; формализована она была в десятом рассказе.

1°. Формализация.

$$(\text{з}_6)\qquad f_0(x)=x\sqrt{1-x^2}\to\max,\quad 0\leqslant x\leqslant 1.$$

2°. Необходимое условие: $f_0'(x)=0$.

3°. Нахождение критических точек. Стационарные точки: $f_0'(x)=0$, т. е.

$$f_0'(x)=(x\sqrt{1-x^2})'=\sqrt{1-x^2}+x(\sqrt{1-x^2})'=$$
$$=\sqrt{1-x^2}-\frac{x^2}{\sqrt{1-x^2}}=0 \iff 2x^2=1 \implies x=\sqrt{2}/2.$$

Таким образом, имеются 3 критические точки: 0, 1, $\sqrt{2}/2$.

4°. Функция f_0 непрерывна на $[0,1]$ и дифференцируема в интервале $(0,1)$. Значит, решение находится среди критических точек. Перебор критических точек: $f_0(0)=f_0(1)=0$, $f_0(\sqrt{2}/2)=1/2$. Таким образом, решение (з$_6$) есть $\sqrt{2}/2$. В этом случае, как это следует из формализации, прямоугольник является квадратом. Ответ: *максимальный из вписанных в круг прямоугольников — квадрат*.

7. Задача Кеплера о вписанном цилиндре. Об этой задаче речь шла в шестом рассказе.

1°. Формализация. Пусть шар имеет радиус R. Обозначим половину высоты цилиндра через x. При этом $0 \leqslant x \leqslant R$. Тогда радиус основания цилиндра равен $\sqrt{R^2 - x^2}$ и его объем равен $\pi r^2 x = 2\pi(R^2 - x^2)x$. Таким образом, приходим к формализации:

$$(з_7) \qquad f_0(x) = 2\pi(R^2 - x^2)x \to \max, \quad 0 \leqslant x \leqslant R.$$

(Фактически эта задача была формализована уже в шестом рассказе.)

2°. Необходимое условие: $f_0'(x) = 0$.

3°. Нахождение критических точек. Стационарные точки:

$$f_0'(x) = (2\pi(R^2 - x^2)x)' = 2\pi(R^2 x - x^3)' = \\ = 2\pi(R^2 - 3x^2) = 0 \iff x_1 = R/\sqrt{3}, x_2 = -R/\sqrt{3}.$$

Второй корень не подходит ($x_2 < 0$). Таким образом, имеются 3 критические точки: 0, $R/\sqrt{3}$ и R.

4°. Функция f_0 непрерывна и дифференцируема всюду. Значит, решение находится среди критических точек. Вследствие того, что $f_0(0) = f_0(R) = 0$, получаем, что решение равно $R/\sqrt{3}$, отсюда следует, что радиус максимального цилиндра равен $\sqrt{R^2 - R^2/3} = R\sqrt{\dfrac{2}{3}}$.

Ответ: *Отношение высоты экстремального цилиндра к диаметру основания равно* $\dfrac{1}{\sqrt{2}}$. Именно этот факт и был установлен Кеплером.

Перейдем теперь к решению некоторых алгебраических задач.

8. Задача Тартальи (рассказ пятый).

1°. Формализация. Обозначим через x меньшее из чисел. Тогда $0 \leqslant x \leqslant 4$ и $8 - x$ — большее число. Их разность равна $8 - 2x$. В итоге

$$(з_8) \qquad f_0(x) = x(8 - x)(8 - 2x) \to \max, \quad 0 \leqslant x \leqslant 4.$$

2°. Необходимое условие: $f_0'(x) = 0$.

3°. Нахождение критических точек. Стационарные точки:

$$f_0'(x) = (x(8 - x)(8 - 2x))' = (2x^3 - 24x^2 + 64x)' = \\ = 6x^2 - 48x + 64 = 0 \iff x_1 = 4 - 4/\sqrt{3}, x_2 = 4 + 4/\sqrt{3}$$

Второй корень не подходит ($x_2 > 4$). Таким образом, имеются три критические точки: 0, 4 и $4 - 4/\sqrt{3}$.

4°. Функция f_0 непрерывна и дифференцируема всюду и рассматривается на конечном отрезке. Значит, решение находится среди критических точек. Но $f_0(0) = f_0(4) = 0$, $f_0(4 - 4/\sqrt{3}) > 0$. Значит, решение (з$_8$) есть $4 - 4/\sqrt{3}$.

Ответ: *бо́льшее число равно* $4 + 4/\sqrt{3}$, *меньшее равно* $4 - 4/\sqrt{3}$. Именно этот факт и был установлен Тартальей.

9. Неравенство между средним геометрическим и средним арифметическим (рассказ пятый).

Рассмотрим вспомогательную экстремальную задачу:

$$(\text{з}_9)\qquad \begin{aligned} &f_0(x) = x_1 \cdot x_2 \cdot \ldots \cdot x_n \to \max,\\ &f_1(x) = x_1 + x_2 + \ldots + x_n = 1,\\ &f_i(x) = x_{i-1} \geqslant 0, \quad i = 2, 3, \ldots, n+1 \quad (x = (x_1, \ldots, x_n)). \end{aligned}$$

Функции f_i непрерывны и их частные производные непрерывны. Совокупность допустимых точек ограничена, ибо $0 \leqslant x_k \leqslant 1$ для любого k. Значит, по теореме Вейерштрасса решение $\widehat{x} = (\widehat{x}_1, \ldots$ $\ldots, \widehat{x}_n)$ существует. При этом, конечно, $\widehat{x}_k \neq 0$, ибо иначе $f_0(\widehat{x}) =$ $= 0$, в то время как существуют допустимые элементы, для которых $f_0(x) > 0$.

Решение $\widehat{x}$ будет, конечно, и локальным максимумом в задаче (з$_9$), откуда (вследствие того, что по доказанному $\widehat{x}_k > 0$) оно будет локальным максимумом и в задаче без неравенств.

1°. Формализация:

$$(\text{з}') \qquad f_0(x) \to \max, \quad f_1(x) = 1.$$

Функция Лагранжа для (з$'$) $\mathscr{L}(x, \lambda_0, \lambda_1) = \lambda_0 f_0(x) + \lambda_1 f_1(x)$.

2°. Необходимое условие — правило множителей Лагранжа:

$$\frac{\partial \mathscr{L}}{\partial x_k} = 0, \quad k = 1, \ldots, n.$$

3°. Нахождение стационарных точек. Обозначим произведение $\widehat{x}_1 \ldots \widehat{x}_n$ через A. Тогда

$$\frac{\partial \mathscr{L}}{\partial x_k}(\widehat{x}, \lambda_0, \lambda_1) = 0 \implies \frac{\lambda_0 A}{\widehat{x}_k} + \lambda_1 = 0.$$

Если допустить, что $\lambda_1 = 0$, то сразу приходим в противоречие с тем, что оба множителя, λ_0 и λ_1, не могут быть нулями. Тогда получаем $\widehat{x}_k = -\lambda_0 A/\lambda_1$, т. е. $\widehat{x}_1 = \ldots = \widehat{x}_n = 1/n$ (из-за того, что $\widehat{x}_1 + \ldots + \widehat{x}_n = 1$).

4°. Вследствие того, что стационарная точка в (з′) единственна, она и доставляет решение задачи.

Теперь докажем нужное неравенство. Пусть $a_1, \ldots, a_n$ — любые неотрицательные числа. Положим $S = a_1 + \ldots + a_n$ и $x_j = a_j/S$. Тогда $x_1 + \ldots + x_n = 1$, и, значит, по доказанному

$$\frac{a_1 a_2 \ldots a_n}{S^n} = x_1 \ldots x_n \leqslant \frac{1}{n^n} \implies a_1 a_2 \ldots a_n \leqslant \left(\frac{a_1 + a_2 + \ldots + a_n}{n}\right)^n,$$

что и требовалось доказать.

10. Неравенства между средним арифметическим и средним квадратическим (рассказ пятый).

Снова рассмотрим вспомогательную задачу.

1°. Формализация:

$$(\text{з}_{10}) \qquad f_0(x) = x_1 + \ldots + x_n \to \max,$$
$$f_1(x) = x_1^2 + \ldots + x_n^2 = 1 \quad (x = (x_1, \ldots, x_n)).$$

Функции f_0 и f_1 и их частные производные непрерывны. Совокупность допустимых точек ограничена, ибо $-1 \leqslant x_k \leqslant 1$, $k = 1, \ldots, n$. Значит, решение $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ существует и можно применять принцип Лагранжа. Функция Лагранжа $\mathscr{L}(x, \lambda_0, \lambda_1) = \lambda_0 f_0(x) + \lambda_1 f_1(x)$.

2°. Необходимое условие:

$$\frac{\partial \mathscr{L}}{\partial x_j} = 0, \quad j = 1, \ldots, n.$$

3°. Нахождение стационарных точек:

$$\frac{\partial \mathscr{L}}{\partial x_j}(\widehat{x}, \lambda_0, \lambda_1) = \lambda_0 + 2\lambda_1 \widehat{x}_j = 0.$$

Если допустить, что $\lambda_1 = 0$, то сразу приходим в противоречие с тем, что оба множителя Лагранжа не могут быть нулями. Тогда получаем, что $\widehat{x}_j = -\lambda_0/2\lambda_1$, т.е. $\widehat{x}_1 = \ldots = \widehat{x}_n = 1/\sqrt{n}$ (из-за того, что $\widehat{x}_1^2 + \ldots + \widehat{x}_n^2 = 1$).

4°. Вследствие того, что стационарная точка единственна, она и доставляет решение задачи.

Теперь докажем нужное неравенство. Пусть $a_1, \ldots, a_n$ — любые числа. Положим $S = (a_1^2 + \ldots + a_n^2)^{1/2}$ и $x_j = a_j/S$. Тогда $x_1^2 + \ldots + x_n^2 = 1$ и, значит, по доказанному

$$\frac{a_1 + \ldots + a_n}{S} = x_1 + \ldots + x_n \leqslant \frac{n}{\sqrt{n}} = \sqrt{n} \implies$$
$$\implies \frac{a_1 + \ldots + a_n}{n} \leqslant \frac{S}{\sqrt{n}} = \left(\frac{a_1^2 + \ldots + a_n^2}{n}\right)^{1/2},$$

что и требовалось. Заодно мы доказали и неравенство (для неотрицательных a_i)

$$\sqrt[n]{a_1 \ldots a_n} \leqslant \left(\frac{a_1^2 + \ldots + a_n^2}{n}\right)^{1/2},$$

решив тем самым планиметрическую и одну из стереометрических задач Кеплера, о которых шла речь в пятом рассказе.

Отметим еще, что, действуя совершенно аналогично, можно было бы доказать следующие общие неравенства для средних. Пусть $a_1, \ldots, a_n$ — неотрицательные числа. Положим

$$S_p = \left(\frac{a_1^p + \ldots + a_n^p}{n}\right)^{1/p}, \quad \Sigma_p = (a_1^p + \ldots + a_n^p).$$

Тогда

$$S_p \leqslant S_q, \quad \text{если} \quad p \leqslant q, \tag{1}$$

$$\Sigma_p \leqslant \Sigma_q, \quad \text{если} \quad p \geqslant q. \tag{2}$$

Выше мы доказали частный случай (1), когда $p = 1$, $q = 2$.

11. Неравенства Коши — Буняковского и Гельдера (рассказ пятый). Пусть $a_1, \ldots, a_n$ — фиксированные числа, не равные одновременно нулю.

1°. Рассмотрим экстремальную задачу:

$$(з_{11}) \qquad f_0(x) = a_1x_1 + \ldots + a_nx_n \to \max,$$

$$f_1(x) = x_1^2 + \ldots + x_n^2 = B^2 \quad (x = (x_1, \ldots, x_n)).$$

Функции f_0 и f_1 и их частные производные непрерывны. Совокупность допустимых точек ограничена, ибо $-B \leqslant x_j \leqslant B$, $j = 1, \ldots, n$. Значит, решение $\widehat{x} = (\widehat{x}_1, \ldots, \widehat{x}_n)$ существует и можно применять принцип Лагранжа. Функция Лагранжа $\mathscr{L}(x, \lambda_0, \lambda_1) = \lambda_0 f_0(x) + \lambda_1 f_1(x)$. Положим $A = (a_1^2 + \ldots + a_n^2)^{1/2}$.

2°. Необходимое условие:

$$\frac{\partial \mathscr{L}}{\partial x_j} = 0.$$

3°. Нахождение стационарных точек:

$$\frac{\partial \mathscr{L}}{\partial x_j}(\widehat{x}, \lambda_0, \lambda_1) = \lambda_0 a_j + 2\lambda_1 \widehat{x}_j = 0, \quad j = 1, \ldots, n.$$

Если допустить, что $\lambda_1 = 0$, то сразу приходим к противоречию с тем, что оба множителя Лагранжа не могут быть нулями. Тогда мы получаем

$$\widehat{x}_j = -\lambda_0 a_j / 2\lambda_1 = Ca_j, \quad j = 1, \ldots, n,$$

откуда вследствие равенства $\widehat{x}_1^2+\ldots+\widehat{x}_n^2=B^2$ получаем, что

$$C^2(a_1^2+\ldots+a_n^2)=B^2 \implies C=\pm\frac{B}{A}.$$

4°. Имеются лишь две стационарные точки. Решением является та точка, где взят знак плюс: $\widehat{x}_j=Ba_j/A$.

Пусть теперь $b_1,\ldots,b_n$ — также любой набор чисел и $B^2=$ $=b_1^2+\ldots+b_n^2$. Тогда по доказанному

$$a_1b_1+\ldots+a_nb_n \leqslant a_1\frac{Ba_1}{A}+\ldots+a_n\frac{Ba_n}{A}=$$
$$=AB=(a_1^2+\ldots+a_n^2)^{1/2}(b_1^2+\ldots+b_n^2)^{1/2}.$$

Но это и есть *неравенство Коши—Буняковского*.

Неравенство Гельдера доказывается точно так же. Укажем без комментариев необходимые выкладки.

1°. Формализация:

$$(з_{11}')\qquad f_0(x)=a_1x_1+\ldots+a_nx_n\to\max,$$
$$f_1(x)=|x_1|^p+\ldots+|x_n|^p=B^p\quad (a_i>0,\ x=(x_1,\ldots,x_n)).$$

Функция Лагранжа: $\mathscr{L}=\lambda_0f_0(x)+\lambda_1f_1(x)$. По определению $A=$ $=(a_1^{p'}+\ldots+a_n^{p'})^{1/p'}$, $p'^{-1}+p^{-1}=1$.

2°. Необходимое условие:

$$\frac{\partial\mathscr{L}}{\partial x_j}=0,\quad j=1,\ldots,n.$$

3°. Нахождение стационарных точек:

$$\lambda_0a_j+p\lambda_1|x_j|^{p-1}\operatorname{sign}x_j=0 \implies \widehat{x}_j=Ca_j^{p'-1},\quad C=\pm\frac{B}{A^{p'/p}}.$$

4°. Решение: $\widehat{x}_j=Ba_j^{p'-1}/A^{p'/p}$, $j=1,\ldots,n$.

Пусть теперь $b_1,\ldots,b_n$ — любой набор неотрицательных чисел. Тогда по доказанному

$$a_1b_1+\ldots+a_nb_n\leqslant a_1\frac{B}{A^{p'/p}}a_1^{p'-1}+\ldots+a_n\frac{B}{A^{p'/p}}a_n^{p'-1}=BA^{p'(1-1/p)}.$$

Но это и есть *неравенство Гельдера*.

И снова вернемся к геометрии.

12. Задача Штейнера (рассказ четвертый). Пусть координаты трех заданных точек таковы: $A=(a_1,a_2)$, $B=(b_1,b_2)$, $C=(c_1,c_2)$. Координату точки D обозначим через (x_1,x_2). Тогда

сумма расстояний от D до A, B и C равна

$$f_0(x_1, x_2) = \sqrt{(x_1 - a_1)^2 + (x_2 - a_2)^2} + \sqrt{(x_1 - b_1)^2 + (x_2 - b_2)^2} + \\ + \sqrt{(x_1 - c_1)^2 + (x_2 - c_2)^2}.$$

Это приводит к задаче без ограничений.

1°. Формализация:

$$(з_{12}) \qquad f_0(x_1, x_2) \to \min.$$

Заметим, что если $x_1^2 + x_2^2$ велико, это означает, что точка D расположена далеко от точек A, B и C, и, значит, сумма расстояний от нее до этих точек также велика. Следовательно, $f_0(x) \to \infty$, когда $x_1^2 + x_2^2 \to \infty$, поэтому можно применить следствие из теоремы Вейерштрасса из предыдущего рассказа. Согласно этому следствию решение $\widehat{x} = (\widehat{x}_1, \widehat{x}_2)$ задачи $(з_{12})$ существует.

Легко понять, что во всех точках x, кроме точек A, B и C, частные производные функции f_0 существуют и непрерывны.

2°. Необходимое (и в силу выпуклости f_0 — достаточное) условие — теорема Ферма. Если $\widehat{x} \neq A, B$ или C, то

$$\frac{\partial f_0(\widehat{x})}{\partial x_1} = \frac{\partial f_0(\widehat{x})}{\partial x_2} = 0.$$

3°. Нахождение стационарных точек:

$$\frac{\partial f_0}{\partial x_1}(\widehat{x}_1, \widehat{x}_2) = \frac{\widehat{x}_1 - a_1}{|\overrightarrow{\widehat{D}A}|} + \frac{\widehat{x}_1 - b_1}{|\overrightarrow{\widehat{D}B}|} + \frac{\widehat{x}_1 - c_1}{|\overrightarrow{\widehat{D}C}|} = 0,$$

$$\frac{\partial f_0}{\partial x_2}(\widehat{x}_1, \widehat{x}_2) = \frac{\widehat{x}_2 - a_2}{|\overrightarrow{\widehat{D}A}|} + \frac{\widehat{x}_2 - b_2}{|\overrightarrow{\widehat{D}B}|} + \frac{\widehat{x}_2 - c_2}{|\overrightarrow{\widehat{D}C}|} = 0.$$

4°. Раскроем геометрический смысл написанных соотношений. Они означают, что три вектора единичной длины

$$e_1 = \frac{\overrightarrow{DA}}{|\overrightarrow{DA}|}, \quad e_2 = \frac{\overrightarrow{DB}}{|\overrightarrow{DB}|}, \quad e_3 = \frac{\overrightarrow{DC}}{|\overrightarrow{DA}|}$$

в сумме равны нулю. Но тогда из них можно составить равносторонний треугольник, т.е. величины углов $A\widehat{D}B$, $B\widehat{D}C$ и $C\widehat{D}A$ равны 120°. Значит, *если решение не совпадает с одной из вершин треугольника ABC, то $\widehat{D}$ — это такая точка, из которой все стороны видны под углом* 120°. Это не что иное, как точка Торричелли, о которой говорилось в четвертом рассказе (там же мы учились ее строить).

Если тупой угол в треугольнике $\geqslant 120°$, то нет такой точки, из которой все стороны видны под углом 120°. Значит, точка $\widehat{x}$

обязана совпадать с одной из вершин, а именно с вершиной тупого угла, ибо против большего угла лежит большая сторона. (Пусть тупой угол — при вершине C. Тогда при естественных обозначениях сторон $c > a$ и $c > b$, откуда следует, что $a+b < a+c$ и $a + b < b + c$, т. е. минимальная сумма — это $a + b$.)

Ответ. *Если все углы треугольника* $< 120°$, *то искомая точка есть точка Торричелли, если один из углов* $\geqslant 120°$, *то искомая точка расположена в вершине этого угла.*

Именно к такому ответу мы пришли в четвертом рассказе.

В первом рассказе была сформулирована задача 4, непосредственно примыкающая к задаче Штейнера. Ответ к ней был приведен в четвертом рассказе. В части, касающейся выпуклого четырехугольника, ответ немедленно следует из достаточности теоремы Ферма для выпуклой задачи без ограничений. В невыпуклом случае, так же, как и в задаче Штейнера, доказывается, что решение существует, и после этого остается проверить, что если точка отлична от вершины, то необходимое условие не может быть выполнено.

13. Задача о наименьшем периметре (четвертый рассказ).

Обратимся к рис. 18. Проведем через точку M прямые, параллельные AB и AC, и точки пересечения с AC и AB соответственно обозначим через N и P. Обозначим $|AN| = a$, $|AP|$ через b, угол BAC через α, отрезок, проходящий через точку M, через $[DE]$ (D лежит на AB, E — на AC), $|NE|$ через x, $|PD|$ через y, углы ADE и DEC соответственно через ψ и φ.

Из подобия треугольников PDM и NME получаем $x/b = a/y \implies xy = ab$. По теореме косинусов

$$|DM| = \sqrt{y^2 + a^2 - 2ya\cos\alpha}, \quad |EM| = \sqrt{x^2 + b^2 - 2xb\cos\alpha},$$

откуда периметр треугольника ADE оказывается равным

$$a + x + \sqrt{y^2 + a^2 - 2ya\cos\alpha} + b + y + \sqrt{x^2 + b^2 - 2xb\cos\alpha}.$$

$1°$. В итоге приходим к такой формализации:

(з$_{13}$)

$$f_0(x, y) = x + \sqrt{x^2 + b^2 - 2xb\cos\alpha} + y + \sqrt{y^2 + a^2 - 2ya\cos\alpha} \to \min,$$
$$f_1(x, y) = xy - ab = 0, \quad x > 0.$$

Выразив y через x и подставив в минимизируемую функцию, приходим к задаче

$$(\text{з}'_{13}), \qquad f_0(x) \to \min, \quad x > 0,$$

где $f_0(x) \to \infty$ при $x \to 0$ и $f_0(x) \to \infty$ при $x \to \infty$ (проверьте).

Значит, по следствию из теоремы Вейерштрасса из одиннадцатого рассказа решение задачи $(\text{з}'_{13})$ — а следовательно, и решение задачи (з_{13}) — существует. Обозначим его $(\widehat{x}, \widehat{y})$. Применяем принцип Лагранжа. Функция Лагранжа $\mathscr{L}(x, y, \lambda_0, \lambda_1) = \lambda_0 f_0(x, y) + \lambda_1 f_1(x, y)$.

2°. Необходимое условие:

$$\frac{\partial \mathscr{L}}{\partial x} = 0, \quad \frac{\partial \mathscr{L}}{\partial y} = 0.$$

3°. Нахождение стационарных точек:

$$\frac{\partial \mathscr{L}}{\partial x} = 0 \implies 1 - \frac{b\cos\alpha - x}{\sqrt{x^2 + b^2 - 2xb\cos\alpha}} + \lambda y = 0,$$

$$\frac{\partial \mathscr{L}}{\partial y} = 0 \implies 1 + \frac{y - a\cos\alpha}{\sqrt{y^2 + a^2 - 2ya\cos\alpha}} + \lambda x = 0.$$

4°. Раскроем геометрический смысл написанных соотношений. Для этого опустим перпендикуляры MR и MS на AC и AB соответственно. Тогда, как видно из рис. 18,

$$\frac{b\cos\alpha - x}{\sqrt{x^2 + b^2 - 2xb\cos\alpha}} = -\frac{|RE|}{|ME|} = \cos\varphi,$$

$$\frac{y - a\cos\alpha}{\sqrt{y^2 + a^2 - 2ya\cos\alpha}} = -\frac{|SD|}{|MD|} = \cos\psi.$$

Соотношения п. 3° можно преобразовать, умножив первое на x, второе на y и использовав написанные равенства:

$$x(1 - \cos\varphi) = y(1 + \cos\psi).$$

Теперь из треугольников MNE и DPM получаем по теореме синусов

$$\frac{|ME|}{\sin\alpha} = \frac{x}{\sin\psi}, \quad \frac{|MD|}{\sin\alpha} = \frac{y}{\sin\varphi}.$$

Откуда окончательно

$$\frac{|ME|(1 - \cos\varphi)}{\sin\varphi} = \frac{|MD|(1 + \cos\psi)}{\sin\psi} \implies |ME| \operatorname{tg}\frac{\varphi}{2} = |MD| \operatorname{tg}\left(\frac{\pi - \psi}{2}\right).$$

Это последнее соотношение имеет следующий геометрический смысл: перпендикуляр к отрезку DE в точке M пересекается с биссектрисами внешних углов D и E в одной точке O. Иначе говоря, вневписанная по отношению к треугольнику ADE окружность проходит через точку M.

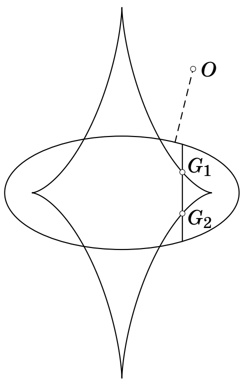
Именно этот ответ и был получен в четвертом рассказе.

14. Задача Аполлония.

В четвертом рассказе говорилось, что геометрические задачи на экстремум встречаются у всех трех величайших математиков античности — Евклида, Архимеда и Аполлония. Но затем были приведены лишь задачи Евклида и Архимеда.

Задачу Аполлония мы не решились привести. И вот почему.

«Коника», или «Конические сечения», — так называется величайшее творение Аполлония (III—II века до н. э.). Принято считать, что «Коника» — это вершина античной математики. К нашей теме относится пятая книга «Конических сечений». Она не переведена на русский язык, и мы приведем цитату из книги «Пробуждающаяся наука» (М.: Гостехиздат, 1954), принадлежащей перу Ван дер Вардена, известного математика и историка математики. Ван дер Варден пишет так: Аполлоний ставит «задачу о том, как провести из одной точки O к коническому сечению самый длинный и самый короткий прямолинейные отрезки. Однако он дает больше, чем обещает: он определяет все проходящие через O прямые, которые пересекают коническое сечение под прямым углом (в настоящее время их называют нормалями), разбирает, при каком положении O задача имеет два, три или четыре решения». Перемещая точку O, он «определяет ординаты граничных точек G_1 и G_2, где число проходящих через O нормалей разом переходит с 2 на 4 и обратно» (рис. 49).



Я не формулировал задачу Аполлония в четвертом рассказе, во-первых, потому, что понятие «коническое сечение» не проходят в школе, а во-вторых, потому, что совершенно не представляю себе, как можно решить задачу, о которой говорит Ван дер Варден, оставаясь в рамках старой элементарной математики. А без такого решения мне не хотелось затрагивать эту задачу в первой части.

В седьмом рассказе, когда речь шла о брахистохроне, отмечалось, что в античные времена рассматривали лишь прямые, окружности и конические сечения (которые, собственно, исследовал Аполлоний). Но кривая, разделяющая области, где число нормалей равно двум и четырем (она называется *астроидой*), принадлежит к совсем другому типу кривых. Она появилась лишь в XVII веке. Трудно даже вообразить, как можно было выразить ее, не прибегая к языку алгебры. (Вспомните эти слова, когда мы найдем уравнение астроиды!)

А теперь перейдем к решению поставленных задач. Сформулируем их так:

1) Как найти расстояние от точки до конического сечения?

2) Сколько нормалей можно провести из точки к коническому сечению?

Мы не будем решать задачи для всех конических сечений, а ограничимся лишь случаем эллипса.

В декартовой системе координат эллипс записывается уравнением $(x_1/a_1)^2 + (x_2/a_2)^2 = 1$. Далее считаем, что $a_1 \geqslant a_2 > 0$, т. е. «ширина» эллипса не меньше его «высоты». Если $a_1 = a_2$, т. е. если ширина и высота совпадают, то эллипс становится окружностью. Решим первую задачу.

1°. Формализация. Пусть точка O имеет координаты (ξ_1, ξ_2). Расстояние от точки O до точки с координатами (x_1, x_2) равно $((x_1 - \xi_1)^2 + (x_2 - \xi_2)^2)^{1/2}$. Для удобства вместо расстояния будем минимизировать квадрат расстояния. В итоге получаем задачу

$$f_0(x_1, x_2) = (x_1 - \xi_1)^2 + (x_2 - \xi_2)^2 \to \min,$$
$$f_1(x_1, x_2) = \left(\frac{x_1}{a_1}\right)^2 + \left(\frac{x_2}{a_2}\right)^2 - 1 = 0.$$

Функции f_0 и f_1 и их частные производные непрерывны, совокупность допустимых точек ограничена, ибо $-a_1 \leqslant x_j \leqslant a_1$, $j = 1, 2$. Значит, решение $\widehat{x} = (\widehat{x}_1, \widehat{x}_2)$ существует и можно применять принцип Лагранжа. Функция Лагранжа:

$$\mathscr{L} = \lambda_0 f_0(x) + \lambda_1 f_1(x).$$

2°. Необходимое условие:

$$\frac{\partial \mathscr{L}}{\partial x_j} = 0, \quad j = 1, 2.$$

3°. Нахождение стационарных точек:

$$\frac{\partial \mathscr{L}}{\partial x_1} = 0 \implies \lambda_0(\widehat{x}_1 - \xi_1) + \frac{\lambda_1 \widehat{x}_1}{a_1^2} = 0,$$
$$\frac{\partial \mathscr{L}}{\partial x_2} = 0 \implies \lambda_0(\widehat{x}_2 - \xi_2) + \frac{\lambda_1 \widehat{x}_2}{a_2^2} = 0.$$

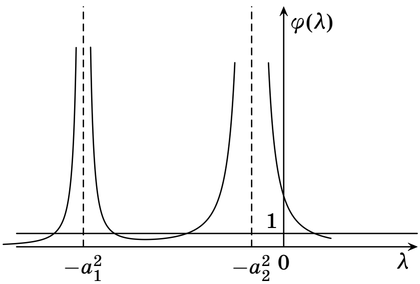
Если допустить, что $\lambda_0 = 0$, то $\lambda_1 \neq 0$ (ведь множители Лагранжа не могут равняться нулю одновременно). Значит, из написанных уравнений последует, что $\widehat{x}_1 = \widehat{x}_2 = 0$, т. е. $f_1(0,0) = 0 \neq -1 = f_1(\widehat{x}_1, \widehat{x}_2)$ — противоречие. Итак, $\lambda_0 \neq 0$, и можно положить $\lambda_0 = 1$. Обозначим при этом $\lambda_1 = \lambda$. Тогда из полученных уравнений

$$\widehat{x}_j - \xi_j + \frac{\lambda \widehat{x}_j}{a_j^2} = 0 \implies \widehat{x}_j = \frac{\xi_j a_j^2}{a_j^2 + \lambda}, \quad j = 1, 2.$$

Подставляя эти соотношения в уравнения эллипса, приходим к уравнению

$$\varphi(\lambda) = \frac{\xi_1^2 a_1^2}{(a_1^2 + \lambda)^2} + \frac{\xi_2^2 a_2^2}{(a_2^2 + \lambda)^2} = 1.$$

4°. Число стационарных точек задачи (т. е. точек, соответствующих тем λ, которые удовлетворяют уравнению $\varphi(\lambda) = 1$) не больше четырех (ибо получилось уравнение четвертой степени, см. также рис. 50; неравенство $\varphi(0) = \xi_1^2/a_1^2 + \xi_2^2/a_2^2 > 1$ на рис. 50 показывает, что функция φ изображена для точки (ξ_1, ξ_2), лежащей вне эллипса).



По-видимому, невозможно записать решение этого уравнения в сколько-нибудь простом явном виде. Но сейчас, когда в нашем распоряжении имеются многочисленные вычислительные средства, мы, конечно, можем получить приближенное решение уравнения $\varphi(\lambda) = 1$ очень скоро и с любой точностью. После того как корни λ_i уравнения будут вычислены, нужно будет найти соответствующие точки $(x_1(\lambda_i), x_2(\lambda_i))$, подставить эти значения в f_0 и найти наименьшее из полученных чисел.

Итак, первая задача решена. Полученные нами соотношения $(\widehat{x}_j - \xi_j) + \lambda \widehat{x}_j / a_j^2 = 0$ имеют такой геометрический смысл: вектор $\xi - \widehat{x}$, соединяющий точку O с минимальной точкой на эллипсе, пропорционален вектору-градиенту функции f_1 в точке $\widehat{x}$, т. е. вектор $\xi - \widehat{x}$ лежит на нормали к эллипсу. Этот факт был впервые установлен Аполлонием.

А теперь примемся за вторую задачу. Выведем уравнения той «разделяющей» кривой, которая отделяет область, где можно провести две нормали, от области, где их можно провести четыре. Из рис. 50 легко понять, что это разделение происходит для тех λ, для которых удовлетворяются условия $\varphi(\lambda) = 1$, $\varphi'(\lambda) = 0$ — ведь именно тогда кривая $y = \varphi(\lambda)$ касается прямой $y = 1$. Значит, надо освободиться от λ из соотношений

$$\varphi(\lambda) = \frac{\xi_1^2 a_1^2}{(a_1^2 + \lambda)^2} + \frac{\xi_2^2 a_2^2}{(a_2^2 + \lambda)^2} = 1,$$

$$\varphi'(\lambda) = -\frac{2\xi_1^2 a_1^2}{(a_1^2 + \lambda)^3} - \frac{2\xi_2^2 a_2^2}{(a_2^2 + \lambda)^3} = 0.$$

Из последнего уравнения получаем

$$a_1^2 + \lambda = A(\xi_1 a_1)^{2/3}, \quad a_2^2 + \lambda = -A(\xi_2 a_2)^{2/3},$$

где

$$A = \frac{a_1^2 - a_2^2}{(\xi_1 a_1)^{2/3} + (\xi_2 a_2)^{2/3}}.$$

Подставляя эти соотношения в уравнение $\varphi(\lambda) = 1$, приходим к уравнению разделяющей кривой

$$(\xi_1 a_1)^{2/3} + (\xi_2 a_2)^{2/3} = (a_1^2 - a_2^2)^{2/3}.$$

Это и есть уравнение астроиды, о которой было сказано выше (как его мог получить Аполлоний?). Вне астроиды каждая точка имеет две нормали, внутри нее — четыре (в частности, конечно, — в центре эллипса), а на самой астроиде — три (за исключением вершин, где имеются две нормали).

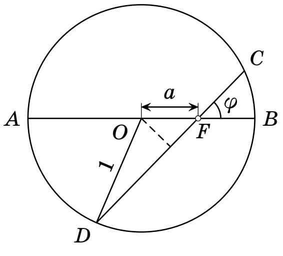
Вот, наконец, мы добрались до того результата, который был получен еще во II веке до н. э.

В 1975 году была организована первая Всесоюзная олимпиада «Студент и научно-технический прогресс». На эту олимпиаду съехалось около ста лучших студентов-математиков всех наших республик. Среди других задач была предложена задача Аполлония: сколько нормалей можно провести из точки к эллипсу? С задачей справился лишь один человек, хотя организаторам, признаться, казалось, что за 22 столетия можно было бы достичь и большего.

Решим еще несколько задач. Летом 1984 года мне предложили провести занятие со школьниками, готовящимися к выступлению на международной математической олимпиаде. Была названа тема занятия — «Математический анализ». Чтобы продемонстрировать силу математического анализа, я решил рассказать то, что уже известно моему читателю из второй части этой книги. Затем на занятии было устроено нечто вроде состязания между анализом и геометрией. Я предлагал задачу, школьники решали ее геометрически, а я — аналитически. Я был убежден в преимуществе анализа и рассчитывал на легкую победу. Однако дело оказалось совсем не таким простым. Мои слушатели были истинными любителями геометрии и, кроме того, — замечательно тренированными в решении задач. Ребятам ничего не стоило придумать неожиданные и очень изящные решения, которые, как мне казалось, совсем нелегко найти. А они считали их тривиальными. Легкой победы не получилось. Но все же нельзя считать, что математический анализ потерпел фиаско. Ниже предлагаются три задачи из этого моего занятия со школьниками, в которых, как мне кажется, «теория» вполне сумела за себя постоять.

Я предлагаю далее лишь аналитические решения. Попробуйте найти «чисто геометрические» решения, которые были бы убедительно проще приведенных!

Первая задача предлагалась на Всесоюзной математической олимпиаде школьников в 1980 году. Ее автор — И. Ф. Шарыгин, известный своим исключительным умением придумывать красивые геометрические задачи (см. [21], задача 349).



15. *Дан круг радиуса единица. На диаметре AB дана точка F, через которую проводится хорда CD. Найти положение хорды, при котором площадь четырехугольника ACBD максимальна.*

Решение. 1°. Формализация. Пусть O — центр круга, $|OF| = a$, угол CFB обозначим через φ (рис. 51). Вспомним, что площадь четырехугольника, вписанного в окружность, равна полупроизведению диагоналей на синус угла между ними. Ясно, что $|CD| = 2\sqrt{1 - a^2\sin^2\varphi}$. Отсюда приходим к такой формализации:

$$\sqrt{1 - a^2\sin^2\varphi}\cdot\sin\varphi \to \max, \quad 0 \leqslant \varphi \leqslant \frac{\pi}{2}.$$

Делаем замену $a\sin\varphi=\sqrt{z}$ и приходим к задаче

$$f(z)=(1-z)z\to\max,\quad 0\leqslant z\leqslant a^2.$$

По теореме Вейерштрасса решение существует.

2°. Необходимое условие — теорема Ферма: $f'(\widehat{z})=0$.

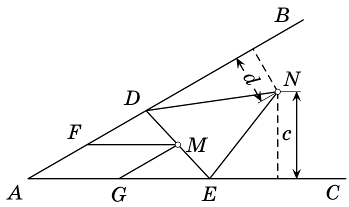
3°. Нахождение критических точек. Стационарная точка одна: $\widehat{z}=1/2$ (если $a^2>1/2$). Критические точки: $\{0,a^2\}$, если $a^2\leqslant 1/2$, и $\{0,1/2,a^2\}$, если $a^2>1/2$.

4°. Сравнивая значения f в критических точках, приходим к ответу: если $0\leqslant a\leqslant 1/\sqrt{2}$, то $\widehat{z}=a^2$, т.е. $\widehat{\varphi}=\pi/2$; если $1/\sqrt{2}<a\leqslant 1$, то $\widehat{z}=1/2$, т.е. $\widehat{\varphi}=\arcsin(1/a\sqrt{2})$.

Эту задачу предложили мне сами школьники. Они были убеждены, что их геометрическое решение будет несравненно проще любого аналитического. Но разве может быть что-нибудь проще решения, приведенного нами?

Следующая задача также принадлежит И. Ф. Шарыгину. Он придумал ее специально для занятий с этой группой. (Ее предполагалось использовать в подготовительных конкурсах, но этого не произошло, так что задача оказалась неизвестной моим слушателям. Они, зная И. Ф. Шарыгина, который также занимался с ними, предвкушали очередную легкую победу геометрии.) Вот как она звучит (см. [21], задача 348).

16. *Дан угол BAC и две точки M и N внутри него. Провести* (*циркулем и линейкой*) *через точку M отрезок DE так, чтобы площадь четырехугольника $ADNE$ была минимальной* (рис. 52)



Решение. 1°. Формализация. Проведем отрезки MF м MG параллельно сторонам AC и AB соответственно и длины этих отрезков обозначим a и b. Опустим из N перпендикуляры на AB и AC и длины этих перпендикуляров обозначим d и c. Тогда удвоенная площадь $ADNE$ равна $(b+x)d+(a+y)c$, где $x=|FD|$, $y=|GE|$. При этом $xy=ab$, что сразу следует из подобия треугольников DFM и MGE. Таким образом,

$$f_0(x,y)=(b+x)d+(a+y)c\to\min,$$
$$f_1(x,y)=xy-ab=0.$$

Из теоремы Вейерштрасса следует (продумайте!), что решение $(\widehat{x},\widehat{y})$ задачи существует (ибо f_0 растет на бесконечности).

Функции f_0 и f_1 и их производные непрерывны, значит, можно применять принцип Лагранжа. Функция Лагранжа

$$\mathscr{L} = \lambda_0 f_0(x) + \lambda_1 f_1(x).$$

2°. Необходимые условия:

$$\frac{\partial \mathscr{L}}{\partial x} = 0, \quad \frac{\partial \mathscr{L}}{\partial y} = 0.$$

3°. Нахождение стационарных точек:

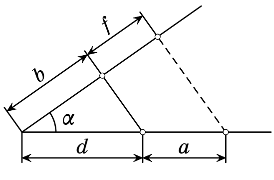
$$\frac{\partial \mathscr{L}}{\partial x} = 0 \implies \lambda_0 d + \lambda_1 \widehat{y} = 0,$$
$$\frac{\partial \mathscr{L}}{\partial y} = 0 \implies \lambda_0 c + \lambda_1 \widehat{x} = 0.$$

Ясно, что $\lambda_0 \neq 0$ (ибо иначе $\widehat{x} = \widehat{y} = 0$, т. е. $\widehat{x}\widehat{y} = 0 \neq ab$), и, значит, можно положить $\lambda_0 = 1$. В итоге приходим к системе уравнений

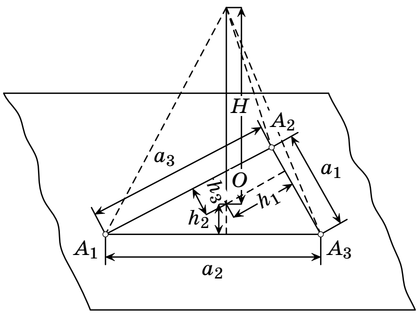
$$xy = ab, \quad \frac{x}{y} = \frac{c}{d}.$$

4°. Построение. Строим (циркулем и линейкой) отрезок f из отношения $f/b = a/d$ (вспомните, как это делается, посмотрев на рис. 53) и затем строим (снова циркулем и линейкой — см. рис. 22) $\widehat{x}$ из соотношения $\widehat{x}^2 = cf \iff \widehat{x} = \sqrt{cf}$.

Существуют и чисто геометрические решения, в частности, очень красивое — И. Ф. Шарыгина. Они приводят к другим методам построения. Найдите их и сравните. Найти более простое геометрическое решение моим слушателям не удалось!



α — любой угол



17. *Среди всех пирамид с данными основанием и высотой найти пирамиду с наименьшей боковой поверхностью.*

Решение. 1°. Формализация (см. рис. 54). Пусть основание пирамиды — это треугольник $A_1A_2A_3$ с длинами сторон a_1, a_2 и a_3 ($a_j > 0$, $j = 1, 2, 3$), O — проекция вершины на плоскость основания. Высоту пирамиды обозначим через H, а через h_1, h_2, h_3 обозначим расстояния от O до прямых, содержащих стороны A_2A_3, A_1A_3 и A_1A_2 соответственно (взятые со знаком плюс, если O лежит в той же полуплоскости, что и треугольник $A_1A_2A_3$, и со знаком минус — если в другой). Тогда имеет место известное (и очевидное) равенство $a_1h_1 + a_2h_2 + a_3h_3 = 2S$, где S — площадь основания. При этом боковая поверхность пирамиды равна $(a_1\sqrt{H^2 + h_1^2} + a_2\sqrt{H^2 + h_2^2} + a_3\sqrt{H^2 + h_3^2})/2$. Приходим к задаче

$$f_0(h_1, h_2, h_3) = a_1\sqrt{H^2 + h_1^2} + a_2\sqrt{H^2 + h_2^2} + a_3\sqrt{H^2 + h_3^2} \to \min$$

$$f_1(h_1, h_2, h_3) = a_1h_1 + a_2h_2 + a_3h_3 - 2S = 0.$$

По теореме Вейерштрасса (продумайте!) решение задачи существует (ибо f_0 растет на бесконечности). Функции f_0 и f_1 и их частные производные непрерывны. Значит, решение $\widehat{h} = (\widehat{h}_1, \widehat{h}_2, \widehat{h}_3)$ существует и можно применять принцип Лагранжа. Функция Лагранжа $\mathscr{L} = \lambda_0 f_0(x) + \lambda_1 f_1(x)$.

2°. Необходимые условия: $\dfrac{\partial \mathscr{L}}{\partial h_j} = 0$ $(j = 1, 2, 3)$.

3°. Нахождение стационарных точек:

$$\frac{\partial \mathscr{L}}{\partial h_j} = 0 \implies \lambda_0 \frac{a_j\widehat{h}_j}{\sqrt{H^2 + \widehat{h}_j^2}} + \lambda_1 a_j = 0, \quad j = 1, 2, 3.$$

Ясно, что $\lambda_0 \neq 0$, и можно считать, что $\lambda_0 = 1$.

4°. Сокращая на a_j в полученных уравнениях, немедленно получаем, что $\widehat{h}_1 = \widehat{h}_2 = \widehat{h}_3$, т. е. *проекция высоты — это центр вписанного круга*.

...После недолгого размышления один из моих слушателей назвал правильный ответ. Я вызвал его к доске, ожидая в очередной раз выслушать «чисто геометрическое» решение, сопровождаемое ироническими замечаниями. Но неожиданно я увидел выписанными функции f_0 и f_1, функцию Лагранжа, ее частные производные и — ответ.

Да здравствует математический анализ! Не так ли?

С нетерпением ожидал я вестей с 25-й Международной математической олимпиады, которая в 1984 году проходила в Праге. Сначала стало известно, что наша команда выступила с успехом, ребята получили пять первых и одну вторую премию и набрали 225 очков. С таким успехом не выступала ни одна команда за всю историю олимпиад. Потом я получил и задачи. Конечно, больше всего меня интересовало, были ли среди них такие, где можно применять методы исследования экстремальных задач. И там-таки была одна такая задача.

18. *Пусть x, y, z — неотрицательные действительные числа такие, что $x+y+z=1$. Докажите, что*

$$0 \leqslant xy + yz + xz - 2xyz \leqslant \frac{7}{27}.$$

Решение. 1°. Формализация:

$$f_0(x,y,z) = xy + yz + xz - 2xyz \to \max \ (\min)$$
$$f_1(x,y,z) = x+y+z-1 = 0,$$
$$x \geqslant 0, \ \ y \geqslant 0, \ \ z \geqslant 0.$$

По теореме Вейерштрасса решение задачи существует и на максимум, и на минимум. Допустим, что некоторое решение $(\widehat{x}, \widehat{y}, \widehat{z})$ таково, что все эти числа отличны от нуля. Но тогда это же решение будет доставлять локальный экстремум задаче $f_0(x,y,z) \to$ $\to \max$ (min), $f_1(x,y,z) = 0$, и к ней можно применять правило множителей Лагранжа.

2°. Необходимое условие (для функции Лагранжа $\mathscr{L} = \lambda_0 f_0(x) - \lambda_1 f_1(x)$):

(i) $$\frac{\partial \mathscr{L}}{\partial x} = 0 \iff \lambda_0(y + z - 2yz) = \lambda_1,$$

(ii) $$\frac{\partial \mathscr{L}}{\partial y} = 0 \iff \lambda_0(x + z - 2xz) = \lambda_1,$$

(iii) $$\frac{\partial \mathscr{L}}{\partial z} = 0 \iff \lambda_0(x + y - 2xy) = \lambda_1.$$

Ясно, что λ_0 не может равняться нулю (иначе $\lambda_1 = 0$). Далее $\lambda_0 = 1$.

3°. Нахождение стационарных точек. Вычитая из уравнения (i) уравнение (ii) (с $\lambda_0 = 1$), получаем

$$y - x - 2z(y - x) = 0 \implies z = \frac{1}{2} \quad \text{или} \quad y = x.$$

Аналогично $y = z$ или $x = 1/2$, и $x = z$ или $y = 1/2$.

4°. Исследование. Если одно из чисел, скажем z, равно $1/2$, то $f_0(x, y, 1/2) \equiv 1/4 < 7/27$. Если же ни одно из чисел не равно $1/2$, то имеется единственная стационарная точка $\widehat{x} = \widehat{y} = \widehat{z} = 1/3$ и $f_0(1/3, 1/3, 1/3) = 7/27$. Наконец, если решение имеет нулевую компоненту, скажем, $\widehat{x}=0$, то $0 \leqslant f_0(0,y,z) = yz \leqslant 1/4$.

О т в е т: максимум $7/27$ достигается при $\widehat{x}=\widehat{y}=\widehat{z}=1/3$, минимум, равный нулю, достигается, скажем, при $\widehat{x}=\widehat{y}=\widehat{z}=1$.

И еще об одной геометрической задаче *). С нею связано воспоминание.

Это было давным-давно, более пятидесяти лет назад. Руководитель кружка, загадочно усмехнувшись, спросил нас: «Один тетраэдр лежит внутри другого. Может ли сумма его ребер оказаться больше, чем сумма ребер объемлющего тетраэдра?» Сначала это показалось нам совершенно невозможным. Ведь тетраэдр лежит внутри другого, как же что-то у него может оказаться больше? Но на самом деле сумма ребер действительно может оказаться большей у меньшего тетраэдра. На втором туре 16-й Всесоюзной математической олимпиады десятиклассникам была предложена такая задача.

19. *Вершины тетраэдра KLMN лежат внутри, на гранях или на ребрах другого тетраэдра ABCD. Докажите, что сумма длин всех ребер тетраэдра KLMN меньше, чем* $4/3$ *суммы длин всех ребер тетраэдра ABCD.*

Это интересный пример, когда можно решить проблему, фактически не приступая к ее стандартному исследованию, ибо слишком многое прояснится сразу после формализации.

Тетраэдр $ABCD$ — это выпуклое замкнутое ограниченное множество. Обозначим его буквой X. В принципе, X можно задать системой из четырех неравенств: $X=\{x=(x_1,x_2,x_3) \mid \langle x,a^i\rangle \leqslant \alpha_i,\ i=1,2,3,4\}$, ($\langle x,y\rangle$ — скалярное произведение x и y).

Наша задача формализуется так:

$$\begin{gathered} f(x^1,x^2,x^3,x^4) = |x^1-x^2|+|x^1-x^3|+|x^1-x^4|+ \\ +|x^2-x^3|+|x^2-x^4|+|x^3-x^4| \to \max, \\ x^i \in X, \quad i=1,2,3,4. \end{gathered}$$

Здесь $x^k=(x^k_1,x^k_2,x^k_3)$, $k=1,2,3,4,$ — точки в трехмерном пространстве, а $|x-y|$ — расстояние от x до y.

Функция f — непрерывная функция $4\times 3=12$ переменных. Тетраэдр X — замкнутое ограниченное множество, задаваемое четырьмя неравенствами. По теореме Вейерштрасса решение

*) Более подробно о ней см. в статье автора, опубликованной в журнале «Квант» 1983, № 1, с. 22—25.

задачи существует. Пусть оно есть $(\widehat{x}^1, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4)$. Из строгой выпуклости f (вытекающей — продумайте — из свойств функции расстояния от фиксированной точки до данной) сразу вытекает, что точки $\widehat{x}^i$ необходимо должны совпадать с вершинами X. Действительно, если, скажем, $\widehat{x}^1$ — не вершина, то существует такой отрезок $[y, z]$, где y и z — точки из X и $\widehat{x}^1 = (y+z)/2$. Но тогда по свойству строго выпуклой функции

$$f(\widehat{x}^1, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4) = f\left(\frac{y+z}{2}, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4\right) < \\ < \frac{1}{2}[f(y, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4) + f(z, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4)].$$

Но отсюда следует, что в одной из точек $(y, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4)$ или $(z, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4)$ функция f имеет большее значение, чем в $f(\widehat{x}^1, \widehat{x}^2, \widehat{x}^3, \widehat{x}^4)$. Получилось противоречие.

А теперь остается лишь несложный перебор. Обозначим через P_{KLMN} периметр тетраэдра $KLMN$, а через P_{ABCD} — периметр тетраэдра $ABCD$.

Если все точки K, L, M и N различны, то тетраэдр $KLMN$ совпадает с $ABCD$, и все ясно, ибо

$$P_{KLMN} = P_{ABCD} < \frac{4}{3}P_{ABCD}.$$

Пусть не более двух вершин тетраэдра $KLMN$ совпадают, скажем, $K = L = A$. Тогда имеются две возможности:

1) две другие вершины тоже совпадают, скажем, $M = N = B$, и тогда из неравенства треугольника

$$P_{KLMN} = 4|AB| = \frac{4}{3}\cdot 3|AB| < \\ < \frac{4}{3}[|AB| + (|AC| + |CB|) + (|AD| + |DB|)] < \frac{4}{3}P_{ABCD};$$

2) вершины M и N различны, скажем, $M = B$, $N = C$; тогда из неравенств треугольника следует:

$$|AB| < |AD| + |DB|, \quad |AC| < |AD| + |DC|,$$

$$|AD| < \frac{1}{3}[|AD| + (|AC| + |CD|) + (|AB| + |BD|)] < \\ < \frac{1}{3}P_{ABCD} \implies P_{KLMN} = 2|AB| + 2|AC| + |BC| < \\ < |AB| + |AD| + |DB| + |AC| + |AD| + |DC| + |BC| = \\ = P_{ABCD} + |AD| < \frac{4}{3}P_{ABCD}.$$

Если же три вершины совпадают, скажем, $K = L = M = A$, $N = B$, то из неравенства п. 1 получаем: $P_{KLMN} = 3|AB| < 4|AB| < \frac{4}{3}P_{ABCD}$. Задача решена.

Нетрудно понять, что число $4/3$ в условии задачи уменьшить нельзя.

Оно достигается для вырожденного тетраэдра $ABCD$, у которого три вершины — A, B и C — совпадают, скажем, с A (и тогда $P_{A'BCD} = 3|AD|$), а у тетраэдра $KLMN$ две вершины совпадают с A, а две — с D (и тогда $P_{KLMN} = 4|AD|$).

Рассказ четырнадцатый

ЧТО БЫЛО ДАЛЬШЕ В ТЕОРИИ ЭКСТРЕМАЛЬНЫХ ЗАДАЧ?

Рассмотрим простейшую задачу о максимуме и минимуме, указывающую нам путь естественного перехода от функций конечного числа переменных к величинам, зависящим от бесконечного числа переменных.

Г. Вольтерра

Излагаемые мною методы не требуют ни построений, ни геометрических или механических рассуждений. Они требуют только алгебраических операций, подчиненных планомерному и однообразному ходу.

Ж. Лагранж

Один старый французский математик сказал: «Математическую теорию только тогда можно считать совершенной, когда ты... берешься изложить ее содержание первому встречному».

Д. Гильберт

1. Об истории математического анализа. Развитие методов решения задач на максимум и минимум неразрывно связано с историей математического анализа. Мы касались этой темы неоднократно. Свяжем же теперь воедино многое из того, о чем рассказывалось раньше. Сначала, как мы помним, задачи на максимум и минимум решались индивидуально, и для каждой из них создавался свой собственный метод. В начале XVII века стала ощущаться потребность отыскать какие-то общие приемы исследования экстремальных задач. Попытки найти алгебраические способы отыскания максимумов и минимумов были сделаны Декартом. Ферма был первым, кто привлек для этих целей идеи, как мы скажем теперь, дифференциального исчисления. Свой метод он открыл, по его собственным словам, еще в 1629 году, но первое достаточно подробное изложение метода содержится в его письмах Робервалю и Декарту, которые он переслал в 1638 году.

«Все учение о нахождении наибольших и наименьших величин основывается на том, что... применяют следующее единственное правило», пишет Ферма и далее излагает суть своей теоремы, о которой говорилось в рассказе десятом.

Можно посоветовать читателю обратиться к книге Декарта «Геометрия» ([8], с. 154), где приведено письмо Ферма, и осознать, как ему удалось выразить свой метод, не прибегая к понятию производной — ведь этого понятия тогда еще не было.

Свои теоретические выводы Ферма подкрепил примером: «*Отрезок AC требуется разделить в E так, чтобы прямоугольник со сторонами AE и EC был наибольшим* (*по площади*)». Нетрудно усмотреть, что фактически это та же самая задача, которую поставил и геометрически решил Евклид в своих «Началах» (рассказ четвертый). Этот пример (причем в формулировке Ферма) был разобран в пятом рассказе.

В начале семидесятых годов XVII столетия Ньютон завершил свой труд под названием «Метод флюксий и бесконечных рядов с приложением его к геометрии кривых». Это сочинение было закончено к 1671 году, но опубликовано лишь в 1736 году. Имеется издание этой работы на русском языке [15].

В ней Ньютон закладывает начала дифференциального и интегрального исчисления, а также теории рядов. Разумеется, Ньютон уделил внимание и нахождению наибольших и наименьших величин. Мимоходом он упоминает прием Ферма, не называя, впрочем, его имени. Он пишет так: «Найди флюксию (т. е. производную) и положи ее равной нулю» (в указ. соч. с. 73). Далее он решает два выразительных примера, где функция задана неявно и в одном из случаев — с иррациональностью. Затем он пишет: «С помощью метода решения этой проблемы можно получить решение следующих проблем» — и указывает девять геометрических задач, которые он умеет исследовать. И снова первая среди них — задача, равносильная евклидовой.

В 1684 году появляется работа Лейбница, в которой также были заложены основы математического анализа. Уже само название ее, которое начинается со слов: «Новый метод нахождения наибольших и наименьших значений...», показывает, какую важную роль сыграла задача о нахождении экстремумов в становлении современной математики. В своей статье Лейбниц не только получает в качестве необходимого условия соотношение $f'(x) = 0$, но и использует второй дифференциал для различения максимума и минимума (впрочем, это также было в то время уже известно Ньютону). С помощью соотношения $f'(x) = 0$ Лейбниц решает ряд конкретных задач, в частности, выводит закон Снеллиуса (рассказ третий).

В трудах Лейбница, значительно опередивших свое время, уже прозвучала мысль о линейной аппроксимации функций, о связи производной и касательной. Эта идея претерпела затем интересную эволюцию, о которой будет рассказано в п. 5.

Исследования Ферма, Ньютона и Лейбница способствовали появлению метода отыскания экстремумов функций одной переменной. Казалось бы, естественно было начать изучение экстремумов функций двух переменных, трех переменных и т. п. Но этого не произошло. История анализа совершила своеобразный зигзаг и сразу стала изучать функции бесконечного числа переменных, и лишь через несколько десятилетий вернулась собственно к конечномерным задачам.

И в задаче Ньютона (рассказ восьмой), и в задаче о брахистохроне (рассказ седьмой), и в классической изопериметрической задаче (рассказ второй) испытанию подвергались «любые кривые». Их нельзя задать с помощью одного, двух, любого конечного числа параметров. В их «произволе» заключено «бесконечно большое число переменных». Недаром именно эти три задачи мы так и не смогли решить в нашем предыдущем рассказе.

С конца XVII века началась разработка теории задач, подобных этим трем. Было создано специальное «исчисление» этих задач. Оно оформилось в XVIII и XIX веках в трудах Эйлера, Лагранжа, Вейерштрасса и других. Его стали называть *вариационным исчислением*.

Несколько позже был развит анализ функций конечного числа переменных. А потом (и сравнительно недавно) было понято, что математический анализ бесконечного числа переменных не является принципиально более сложным, чем конечномерный анализ. И снова мысль о создании бесконечномерного анализа отталкивалась от потребности решать экстремальные задачи (см. слова Г. Вольтерра, поставленные нами в качестве эпиграфа, где Вольтерра имеет в виду классическую изопериметрическую задачу). Коснемся чуть подробнее следующей темы.

2. Что такое функции бесконечного числа переменных? В рассказе девятом мы обсуждали вопрос: что такое функция? Сначала было рассказано о функциях одной переменной, когда одному числу x ставится в соответствие по определенному правилу число y. Затем речь шла о функции двух переменных, когда паре чисел (x_1, x_2) ставится в соответствие (и снова по определенному правилу) число y. Наконец, было сказано и о функциях n переменных. Но еще до девятого рассказа

(и причем неоднократно) нам пришлось столкнуться с функциями бесконечного числа переменных, когда переменными были сами функции. (Вспомним формулу (2) в рассказе седьмом и формулу (8′) в рассказе восьмом.) Такого рода *функционалы* (так зачастую называют в математике функции от функций) изучали почти за двести лет до того момента, как научились обращаться с функциями бесконечного числа переменных так же свободно, как и с функциями одной переменной.

Рассмотрим совокупность (в математике говорят еще — *пространство*) всех непрерывных функций на отрезке $[a, b]$ числовой прямой. Ее обозначают $C[a, b]$. Пусть эти непрерывные функции теперь играют роль аргумента. Попробуем осмыслить этот факт. Вспомним, что такое функция одной переменной, т. е. функция, заданная на числовой прямой $\mathbb{R}$ (скажем, все та же функция $y = y(x) = \sqrt{1 + x^2}$). Это, как мы помним, — правило, дающее возможность по заданному числу x получать число y (в конкретном случае надо взять x, возвести его в квадрат, прибавить единицу и затем извлечь корень.)

Попробуем теперь понять, что такое функция $F(y)$ на пространстве $C[a, b]$. По самому смыслу понятия «функция» это должно быть некоторое правило, которое по непрерывной функции $y(x)$, заданной на $[a, b]$, дает возможность вычислить число $F(y)$.

Возьмем несколько примеров. Для определенности пусть дальше $a = 0$, а $b = 1$.

Пример 1. $F_1(y) = 2y(0)$.

Что здесь предписано? По функции $y(x)$ следует сначала вычислить то число, которое она принимает в точке нуль, а потом умножить его на 2. Вспомним несколько известных нам функций и вычислим для них значение $F_1(y)$. Если $y(x) = x$, то $y(0) = 0$, и, значит, $F_1(y) = 0$; если $y(x) = \sqrt{1 + x^2}$, то $y(0) = 1$ и $F_1(y) = 2$; если $y(x) = 5 \cdot 2^x$, то $y(0) = 5$ и $F_1(y) = 10$.

Представьте себе теперь, что мы играем в такую игру: вам задают подряд несколько функций, а вы вычисляете $F_1(y)$ от каждой. К примеру, вам предлагают $y(x) = 3\cos(x + 2)$, или $5\ln(x + 3)$, или еще что-то. Мне кажется, вы не затруднитесь ответить, чему будет равно $F_1(y)$ в этих случаях. Понять правила этой игры — это и значит понять, что такое функция F_1.

Пример 2. $F_2(y) = \int\limits_0^1 y(x)\,dx$. (Этот функционал выражает не что иное, как *площадь под графиком* $y(x)$.) И снова давайте играть в ту же игру. Я вам даю $y(x)=1$ — вы мне вычисляете: $F_2(y)=1$; я вам — $y(x)=x$, вы мне — $F_2(y)=1/2$; я вам — $y(x)=\sin x$, вы мне — $F_2(y)=1-\cos 1$ и т. д. Так мы познакомились с очень важным функционалом, *функционалом площади.*

Но можно придумать что-нибудь похитрее.

Пример 3. $F_3(y)=(2y(0))^3-\left(\int\limits_0^1 y(x)\,dx\right)^2.$

Здесь предписание более сложное. Если вам дают какую-то функцию $y(x)$, то вы должны сначала вычислить $y(0)$, затем это число умножить на два, потом полученное число возвести в куб, далее — найти интеграл от $y(x)$ по отрезку $[0,1]$, возвести этот интеграл в квадрат, а затем из «куба» вычесть этот «квадрат». Скажем, если вам задали уравнение $y(x)=x$, то получится $-1/4$, а если $y(x)=5\cdot 2^x$, то и вовсе придется потрудиться. Но важно, что «в принципе» вы выполните поставленную задачу и по каждой функции $y(x)$ (которую можно отчетливо описать) получите число $F_3(y)$.

В рассказе девятом я писал в соответствующем месте: «А теперь давайте определим и изобразим несколько важнейших функций». Здесь что-либо изобразить трудно, ведь нужно рисовать в бесконечномерном пространстве, а «определить» — попробуем.

Простейшая функция — постоянная $F(y)=c$. Каждой непрерывной функции $y(x)$ сопоставляется единственное число c.

Следующие по сложности функции — линейные. Что это значит — «линейная функция» от функции? Это значит, что сумму любых функций она переводит в сумму чисел (т. е. $F(y_1+y_2)=F(y_1)+F(y_2)$) и, кроме того, что $F(ay)=aF(y)$ для любого числа a и любой функции $y(x)$.

Функции из примеров 1 и 2, приведенных нами выше, — линейные, функция из третьего примера линейной не является. Здесь, в бесконечномерном случае, линейных функций очень много. (В каком-то смысле их даже «больше», чем всех непрерывных функций. Ведь если φ — непрерывная функция, то ей можно сопоставить линейный функционал

$$F_\varphi(y)=\int\limits_0^1 \varphi(x)y(x)\,dx,$$

а функционал $F_1(y)=2y(0)$ не может быть представлен в этом виде.)

Бесконечномерный анализ изучает функции от «бесконечного числа переменных», точнее — функционалы на бесконечномерных пространствах (типа пространства $C[a,b]$). Приведем пример еще одного важного бесконечномерного пространства, с которым на протяжении двух веков в основном оперировало вариационное исчисление. Это пространство $C^1[a,b]$ непрерывно дифференцируемых функций $y(x)$, т. е. таких функций $y(x)$, которые сами непрерывны и их производные также непрерывны на отрезке $[a,b]$.

В пространстве $C^1[a,b]$ определены функционалы, имеющие важный геометрический или физический смысл. Приведем примеры.

Пример 4. Функционал «длины»:

$$L(y) = \int_a^b \sqrt{1 + (y'(x))^2}\,dx.$$

Пример 5. Функционал И. Бернулли — «время движения вдоль кривой» (см. формулу (2) в рассказе седьмом):

$$T(y) = \int_0^a \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{2gy(x)}}\,dx.$$

Пример 6. Функционал Ньютона — сопротивление движению в редкой среде (см. формулу (8′) в рассказе восьмом):

$$F(y) = 2K\int_0^R \frac{x\,dx}{1 + (y'(x))^2}.$$

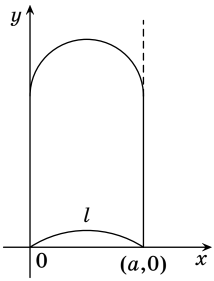
Эти примеры можно неограниченно продолжать. Но не будем делать этого, а перейдем сразу к вопросу о том, *как ставятся экстремальные задачи для функционалов*. В рассказе девятом говорилось: для того, чтобы точно сформулировать экстремальную задачу, необходимо описать максимизируемую или минимизируемую функцию и ограничение. Ограничения, как мы помним, задаются обычно равенствами и неравенствами. В бесконечномерном случае все остается без изменений. Здесь также нужно описать минимизируемый и максимизируемый функционал (а значит, и пространство, на котором он определен) и ограничения.

Формализуем те задачи из первой части, которые мы не решали пока во второй. Во всех случаях будем исследовать задачи в пространстве C^1.

Задача Дидоны. Вспомним историю о Дидоне (начало второго рассказа). Анализируя ситуацию, в которой оказалась финикийская царевна, можно представить (среди прочих) следующие две возможности поставить задачу оптимизации.

А) Первая задача Дидоны, или классическая изопериметрическая задача. *Требуется указать оптимальную форму участка земли, который при заданной длине периметра l имел бы наибольшую площадь.*

Эта задача и исследовалась нами во втором рассказе. Другие постановки получаются, если, как это естественно считать, Дидона хотела сохранить выход к морю. Для простоты рассмотрим случай прямолинейного берега и будем при этом считать, что Дидоне были указаны границы, за пределы которых она не могла выйти (рис. 55). Тогда получается



Б) Вторая задача Дидоны. *Требуется среди всех дуг длины l, лежащих в полуполосе $0 \leqslant x \leqslant a$, $y > 0$ с заданными концами в точках $(0,0)$ и $(a,0)$, найти такую, которая вместе с отрезком $y = 0$, $0 \leqslant x \leqslant a$, ограничила бы фигуру наибольшей площади.*

Мы ограничим себя формализацией только второй задачи *). Пусть $y = y(x)$ — уравнение дуги. Выше нам встречались функционалы «площадь» и «длина». Вспомнив, как они задаются, приходим к такой постановке:

$$S(y) = \int_0^a y(x)dx \to \max, \quad L(y) = \int_0^a \sqrt{1 + (y'(x))^2}\, dx = l$$

при граничных условиях $y(0) = 0$, $y(a) = 0$.

Максимизируемый функционал здесь — это площадь. Ограничение задается равенством $L(y) = l$, где $L(y)$ — функционал длины. Граничные условия также задаются равенствами $\Gamma_1(y) = 0$, $\Gamma_2(y) = 0$, где $\Gamma_1(y) = y(0)$, $\Gamma_2(y) = y(a)$.

Задача о брахистохроне, по сути дела, была уже формализована нами. Вспомнив еще раз формулу (2) в рассказе седьмом, мы получим нужную формализацию:

$$T(y) = \int_0^a \frac{\sqrt{1 + (y'(x))^2}}{\sqrt{2gy(x)}}\, dx \to \min$$

с граничными условиями $y(0) = 0$, $y(a) = b$.

---

*) Формализации первой задачи требуют рассмотрения функционалов от пары функций. Такого рода задачи мы опишем в конце этого рассказа. Решение же первой задачи нетрудно получить из решения второй.

Здесь минимизируемый функционал — это функционал Бернулли.

Задача Ньютона также фактически была формализована нами в восьмом рассказе. Эта формализация выглядит так:

$$F(y) = 2K \int_0^R \frac{x\,dx}{1 + (y'(x))^2} \to \min, \quad y'(x) \geqslant 0,$$

с граничными условиями $y(0) = 0$, $y(R) = H$.

Обратим особое внимание на ограничение: $y'(x) \geqslant 0$ — условие монотонности. Оно встретилось только здесь.

Напомним, что все эти задачи были решены нами по-разному (по крайней мере, внешне). Но во всех решениях присутствовало одно и то же: кривую мы всюду аппроксимировали ломаными и тем самым сводили задачу к конечномерной.

Такой метод решения задач подобного типа был осуществлен Эйлером, а его предшественником был, как мы помним, Лейбниц. Этот метод и его модификации получили название *прямых методов в вариационном исчислении*. Их и до сих пор применяют для численного решения задач вариационного исчисления.

3. Задачи вариационного исчисления и принцип Лагранжа для них. Мы много раз употребляли уже термин «вариационное исчисление», а теперь подошло время, чтобы его точно объяснить. Пусть нам задана некоторая, скажем, непрерывная функция трех переменных $f(x, y, z)$. Рассмотрим следующий функционал:

$$F(y) = \int_a^b f(x, y(x), y'(x))\,dx.$$

Этот функционал можно рассматривать в различных пространствах, но чаще всего его исследуют в пространстве C^1. *Вариационным исчислением называется раздел теории экстремальных задач, где изучают максимумы и минимумы таких функционалов при различного рода ограничениях* (о которых мы немного скажем дальше).

Посмотрим, что надо сделать, чтобы по функции $y(x)$ из $C^1[a, b]$ получить число $F(y)$. Сначала надо продифференцировать $y(x)$. Затем надо подставить $y(x)$ во второй аргумент, а $y'(x)$ — в третий аргумент функции $f(x, y, z)$. Тогда получится функция одной переменной, которая числу x ставит в соответствие число $f(x, y(x), y'(x))$. А в заключение надо взять интеграл от этой функции, и тогда получится искомое число $F(y)$.

Вернемся на короткое время к двенадцатому рассказу. В первом пункте этого рассказа была поставлена задача о минимизации или максимизации функций нескольких переменных при ограничениях типа равенств и неравенств. Если бы неравенств не было, то задача приняла бы такой вид:

$$(\text{з})\qquad F_0(x)\to\min\ (\max),\quad F_i(x)=0,\quad i=1,\ldots,m,$$

где $x=(x_1,\ldots,x_n)$, а $F_i(x)$ — функции n переменных (мы сознательно заменили малые буквы f_i большими F_i).

Давайте теперь изучать точно такие же задачи, только здесь F_i — не функции многих переменных, а введенные выше функционалы типа $F(y)$. Функционалы типа $F(y)$ называют *функционалами классического вариационного исчисления*. Примеры их встретились у нас в предыдущих пунктах. Например, в функционале длины $f(x,y,z)=\sqrt{1+z^2}$, в функционале площади $f(x,y,z)=y$, в функционале И. Бернулли $f(x,y,z)= =\sqrt{1+z^2}/\sqrt{2gy}$ и т. п.

Пусть далее задан набор функций $f_0(x,y,z)$, $f_1(x,y,z)$, ..., ..., $f_m(x,y,z)$. Рассмотрим соответствующие им функционалы вариационного исчисления $F_0(y),F_1(y),\ldots,F_m(y)$ и следующую вариационную задачу:

$$(\text{з}_1)\qquad F_0(y)\to\min\ (\max),\quad F_i(y)=\alpha_i,\quad i=1,\ldots,m,$$

где функции $y(x)$ удовлетворяют краевым условиям на концах отрезка $[a,b]$: $y(a)=y_0$, $y(b)=y_1$.

Ее называют *изопериметрической задачей* классического вариационного исчисления.

Множество функций $y(x)$ на $C^1[a,b]$, удовлетворяющих условиям $F_i(y)=\alpha_i$, $i=1,\ldots,m$, $y(a)=y_0$, $y(b)=y_1$, называют *допустимыми в задаче* (з$_1$).

Если ограничения типа $F_i(y)=\alpha_i$ отсутствуют, то задача принимает вид

$$(\text{з}_2)\quad F_0(y)\to\min\ (\max)\iff\int_a^b f_0(x,y(x),y'(x))\,dx\to\min\ (\max)$$

по всем $y(x)$ таким, что $y(a)=y_0$, $y(b)=y_1$, и задачу (з$_2$) называют *простейшей задачей* классического вариационного исчисления.

Задача о брахистохроне относится к числу простейших, вторая задача Дидоны — к числу изопериметрических (с этим и связан термин «изопериметрический» в применении к задаче (з$_1$)).

Задача Ньютона не относится ни к тому, ни к другому типу, ибо там имеется дополнительное ограничение $y'(x) \geqslant 0$, которого ни в ($з_1$), ни в ($з_2$) нет.

Возникает вопрос, как определить понятие локального минимума (максимума) в задаче ($з_1$). Для этого надо задать какую-то «меру близости» функций из $C^1[a, b]$ друг к другу. В качестве меры близости функции $y(x)$ к функции, тождественно равной нулю, берут число

$$\|y\|_1 = \max_{x\in[a,b]} |y(x)| + \max_{x\in[a,b]} |y'(x)|,$$

(его называют *нормой* функции $y(x)$), а в качестве меры близости функций $y_1(x)$ и $y_2(x)$ — число $\|y_1 - y_2\|_1$. В пространстве $C[a, b]$ норма вводится так:

$$\|y\|_0 = \max_{x\in[a,b]} |y(x)|.$$

Теперь определение локального минимума становится совершенно аналогичным определению 1 из двенадцатого рассказа.

О п р е д е л е н и е. Говорят, что функция $\widehat{y}(x)$ *доставляет локальный минимум* (*максимум*) *в задаче* ($з_1$), если можно указать такое $\varepsilon > 0$, что для всех допустимых в задаче ($з_1$) функций, удовлетворяющих неравенству

$$\|y - \widehat{y}\|_1 < \varepsilon,$$

выполнено неравенство $F_0(y) \geqslant F_0(\widehat{y})$ $(F_0(y) \leqslant F_0(\widehat{y}))$.

Возникает вопрос: как решать задачу ($з_1$)?

Напомним, в чем состоял принцип Лагранжа в применении к задаче (з). Он состоял из двух утверждений.

1) Для задачи без ограничений необходимым условием экстремума в точке $\widehat{x}$ является соотношение

$$F_0'(\widehat{x}) = 0$$

(т е о р е м а Ф е р м а).

2) Для решения задачи (з) надо составить функцию Лагранжа и затем поступить с ней так, как будто переменные независимы (т. е. нужно применить теорему Ферма). Второе утверждение мы назвали п р и н ц и п о м Л а г р а н ж а.

В применении к задаче ($з_1$) все оказывается совершенно аналогичным. Только теорему Ферма требуется переосмыслить. Именно, имеет место следующая теорема.

Т е о р е м а (Эйлера). *Пусть в простейшей задаче* ($з_2$) f_0 — *непрерывно дифференцируемая функция трех переменных. Тогда,* *если функция $\widehat{y}(x)$ доставляет локальный экстремум (минимум или максимум) простейшей задаче* (з$_2$), *то она удовлетворяет следующему уравнению:*

$$\frac{d}{dx} f_{0y'}(x, \widehat{y}(x), \widehat{y}'(x)) - f_{0y}(x, \widehat{y}(x), \widehat{y}'(x)) = 0. \qquad (1)$$

Оно называется *уравнением Эйлера* задачи (з$_2$). Его допустимые решения называются *стационарными* точками задачи (з$_2$), или ее *экстремалями*.

Уравнение (1) и является расшифровкой уравнения типа $F_0'(\widehat{x}) = 0$ в применении к простейшей задаче. Этот факт мы постараемся объяснить в следующем пункте. Отметим, что для решения конкретных задач не требуется понимания, откуда берется уравнение (1). Таким образом, алгоритм (идущий от Эйлера) исследования простейших задач состоит в следующем.

Нужно найти все решения уравнения (1) (*а они зависят от двух параметров*), *которые проходят через данные точки, и выбрать среди них те, для которых функционал F_0 принимает наименьшее* (*наибольшее*) *значение.*

Нетрудно показать непосредственным дифференцированием, что если f_0 не зависит от x, то уравнение (1) допускает следующее соотношение («интеграл»):

$$f_{0y'}(\widehat{y}(x), \widehat{y}'(x)) - y'(x) f_{0y'}(\widehat{y}(x), \widehat{y}'(x)) \equiv \mathrm{const}. \qquad (1')$$

Иначе говоря, любое решение уравнения (1) удовлетворяет соотношению (1$'$).

Что же касается общей задачи (з$_1$), то к ней метод Лагранжа применим безо всяких изменений. Нужно составить функцию Лагранжа

$$\mathscr{L} = \lambda_0 F_0(y) + \lambda_1 F_1(y) + \ldots + \lambda_m F_m(y).$$

Ее можно записать так:

$$\mathscr{L} = \int_a^b f(x, y(x), y'(x))\, dx,$$

где

$$f(x, y, z) = \lambda_0 f_0(x, y, z) + \lambda_1 f_1(x, y, z) + \ldots + \lambda_m f_m(x, y, z).$$

И далее следует поступать так, как будто мы имеем дело с нахождением экстремума функции $\mathscr{L}$, а функции y не связаны. Иначе говоря, надо написать уравнение Эйлера

$$\frac{d}{dx} f_{y'} - f_y = 0 \iff \frac{d}{dx}(\lambda_0 f_{0y'} + \ldots + \lambda_m f_{my'}) - (\lambda_0 f_{0y} + \ldots + \lambda_m f_{my}) = 0.$$

Все сказанное основывается на следующей теореме.

Теорема (правило множителей Лагранжа для изопериметрических задач). *Пусть функции* $f_0, \ldots, f_m$ *непрерывно дифференцируемы. Тогда, если функция* $\widehat{y}(x)$ *достигнет локального экстремума* (*минимума или максимума*) *в задаче* (з$_1$), *найдутся числа* $\lambda_0, \ldots, \lambda_m$, *не равные нулю одновременно и такие, что выполнено уравнение Эйлера*

$$\frac{d}{dx}(\lambda_0 f_{0y'}(x, \widehat{y}(x), \widehat{y}'(x)) + \ldots + \lambda_m f_{my'}(x, \widehat{y}(x), \widehat{y}'(x))) - \\ - (\lambda_0 f_{0y}(x, \widehat{y}(x), \widehat{y}'(x)) + \ldots + \lambda_m f_{my}(x, \widehat{y}(x), \widehat{y}'(x))) = 0. \quad (2)$$

Допустимые решения уравнения (2) называются *стационарными решениями*.

Правило множителей Лагранжа позволяет дать такой рецепт поиска решения задачи (з$_1$). Разбиваем его на те же 4 этапа.

1°. Формализация задачи.

2°. Применение принципа Лагранжа, т. е. составление уравнения (2) совместно с уравнениями $F_i(y) = \alpha_i$, и граничными условиями $y(a) = y_0$, $y(b) = y_1$.

3°. Нахождение всех стационарных решений.

4°. Отбор стационарных решений, являющихся решениями задачи.

Применим этот алгоритм к решению наших двух задач — о брахистохроне и Дидоны.

Решение задачи о брахистохроне.

1°. Задача о брахистохроне была уже формализована нами как простейшая задача с

$$f_0(x, y, z) = \frac{\sqrt{1+z^2}}{\sqrt{2gy}}$$

(не зависит от x).

2°. Необходимое условие — уравнение Эйлера

$$\frac{d}{dx} f_{0y'} - f_{0y} = 0$$

допускает интеграл

$$f_0 - y' f_{0y'} = \text{const} \iff \frac{\sqrt{1+y'^2}}{\sqrt{2gy}} - \frac{y'^2}{\sqrt{1+y'^2}\sqrt{2gy}} = \\ = \text{const} \implies \sqrt{1+y'^2}\sqrt{y} = C, \quad (*)$$

где C — некоторая константа. Вспомним, что именно к этому соотношению и пришел сам И. Бернулли.

3°. Нахождение стационарных точек — это нахождение решений уравнения (∗). Но это уравнение было проинтегрировано нами. При этом оказалось, что решения его — это семейство циклоид.

4°. В седьмом рассказе было показано, что имеется лишь одна допустимая циклоида из построенного в предыдущем пункте семейства. Она и будет решением задачи (это, конечно, еще нужно обосновать).

Решение второй задачи Дидоны.

1°. Задача Дидоны была уже формализована нами как изопериметрическая задача с $f_0(x, y, z) = y$, $f_1(x, y, z) = \sqrt{1+z^2}$.

2°. Принцип Лагранжа. Составляем сумму $f = \lambda_0 f_0 + \lambda_1 f_1$ и выписываем уравнение Эйлера

$$\frac{d}{dx}\lambda_1\frac{y'}{\sqrt{1+y'^2}} - \lambda_0 = 0. \tag{$**$}$$

3°. Решаем полученное уравнение, учитывая, что если $\lambda_0 = 0$, то $\widehat{y} \equiv 0$ (проверьте, учтя граничные условия), что возможно, лишь если $l = a$. Значит, если $l > a$, можно считать, что $\lambda_0 = 1$. Тогда из (∗∗) получаем

$$\frac{d}{dx}\lambda_1\frac{y'}{\sqrt{1+y'^2}} = 1 \implies \frac{\lambda_1 y'}{\sqrt{1+y'^2}} = x + c \implies \frac{y'}{\sqrt{1+y'^2}} = Cx + D \implies$$

$$\implies \frac{y'^2}{1+y'^2} = (Cx+D)^2 \implies \frac{dy}{dx} = \frac{\pm(Cx+D)}{\sqrt{1-(Cx+D)^2}} \implies$$

$$\implies dy = \pm\frac{(Cx+D)\,dx}{\sqrt{1-(Cx+D)^2}} \implies d\left(y \pm \frac{1}{C}\sqrt{1-(Cx+D)^2}\right) = 0 \implies$$

$$\implies (x+a)^2 + (y+b)^2 = r^2.$$

Получилось семейство всех окружностей.

4°. Нетрудно теперь найти нужное решение. Если $a < l \leqslant \pi a/2$, то в построенном семействе окажется единственная окружность длины l, проходящая через точки $(0, 0)$ и $(a, 0)$. Если же $l > \pi a/2$, то решением будет полуокружность с центром в точке $\left(\frac{a}{2}, \frac{1}{2}\left(l - \pi\frac{a}{2}\right)\right)$ радиуса $a/2$, «дополненная» отрезками $x = 0$, $0 \leqslant y \leqslant l - \pi a/2$, $x = a$, $0 \leqslant y \leqslant l - \pi a/2$ (см. рис. 55).

Отметим отличие наших решений в рассказе тринадцатом и здесь.

Там задачи были исследованы до конца. Здесь все-таки остается неопределенность, связанная с существованием решения. Существование решения в задачах вариационного исчисления доказывается труднее, чем в конечномерном случае. Более того, там часто решения просто не существует. Вот например, в только что исследованной задаче Дидоны при $l > \pi a/2$ решения в обычном значении не существует, ибо поднятая полуокружность, «дополненная отрезками», не есть функция $y(x)$, соединяющая точки $(0,0)$ и $(a,0)$.

Здесь, как говорят математики, имеется «обобщенное» решение.

Не так просто, как может показаться, обстоит дело и с брахистохроной, ибо циклоида не непрерывно дифференцируема, т. е., иначе говоря, решение задачи о брахистохроне не существует в той совокупности функций (в пространстве $C^1[0,a]$), в которой рассматривалась сама задача.

Знаменитому математику XX века Давиду Гильберту (его слова были поставлены эпиграфом к этому рассказу, и к ним мы еще вернемся) принадлежит идея о том, что любая естественно поставленная вариационная задача должна иметь решение, «если в случае необходимости самому понятию решения придать расширенное толкование».

И действительно, идея Гильберта оправдывается в большинстве задач, в частности в задаче о брахистохроне, задаче Дидоны, а также в двух задачах, которые были сформулированы в конце седьмого рассказа. К решению последних мы приступаем.

З а д а ч а Л о п и т а л я. 1°. Ф о р м а л и з а ц и я. Время распространения света от точки $(0, y_0)$ до точки (a, y_1) для среды, где скорость распространения зависит лишь от высоты y и равна $v(y)$, выражается интегралом

$$T(y) = \int_0^a \frac{\sqrt{1 + (y'(x))^2}}{v(y)}\, dx.$$

Для того чтобы убедиться в этом, нужно еще раз посмотреть на формулу (2) в рассказе седьмом. В итоге получается следующая простейшая задача классического вариационного исчисления:

$$\int_0^a \frac{\sqrt{1 + (y'(x))^2}}{y}\, dx \to \min.$$

$$y(0) = y_0, \;\; y(a) = y_1 \qquad \left(f_0(x, y, z) = \frac{\sqrt{1 + z^2}}{y} \right).$$

2°. Н е о б х о д и м о е у с л о в и е — уравнение Эйлера. Оно допускает интеграл, который мы и выписываем:

$$f_0 - y' f_{0y'} = \mathrm{const} \iff y\sqrt{1 + y'^2} = D.$$

3°. Найдем стационарные точки:

$$y\sqrt{1+y'^2} = D \implies \frac{y\,dy}{\sqrt{D^2-y^2}} = dx \implies$$

$$\implies x - C_1 = \int \frac{y\,dy}{\sqrt{D^2-y^2}} \implies (x-C_1)^2 + y^2 = D^2.$$

Уравнение Эйлера проинтегрировано. Получили семейство полуокружностей с центрами на оси Ox.

4°. Легко понять, что через любые две точки $(0, y_0)$ и (a, y_1) можно провести одну и только одну окружность нашего семейства. Эта окружность и будет решением задачи, но доказательство этого — за пределами наших возможностей. *Те же самые полуокружности являются прямыми в интерпретации Пуанкаре плоскости Лобачевского.*

Задача о минимальной поверхности вращения. 1°. Формализация:

$$\int_{x_0}^{x_1} y\sqrt{1+(y'(x))^2}\,dx \to \min, \quad y(x_0) = y_0, \quad y(x_1) = y_1.$$

Вспомним: функционал «площадь поверхности вращения» уже выписывался нами в конце седьмого рассказа. Здесь был отброшен лишь множитель 2π. Таким образом, в рассматриваемом случае

$$f_0(x, y, z) = y\sqrt{1+z^2}.$$

2°. Необходимое условие — уравнение Эйлера, допускающее интеграл, который и выписывается:

$$f_0 - y' f_{0y'} = \mathrm{const} \implies \frac{1+y'^2}{y^2} = D^2 \implies \frac{dy}{\sqrt{D^2y^2-1}} = dx.$$

3°. Нахождение стационарных точек (экстремалей). Решим уравнение из п. 2 с помощью подстановки:

$$Dy = \frac{e^t + e^{-t}}{2} \implies D\,dy = \frac{(e^y - e^{-t})dt}{2},$$

$$\sqrt{D^2y^2-1} = \sqrt{\frac{(e^t+e^{-t})^2}{4} - 1} = \frac{e^t - e^{-t}}{2} \implies$$

$$\implies \frac{dy}{\sqrt{D^2y^2-1}} = \frac{dt}{D} \implies \frac{dt}{D} = dx \implies t = Dx + D_1,$$

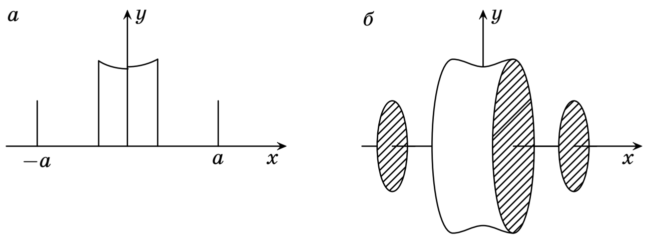
$$y = \frac{e^{Dx+D_1} + e^{-(Dx+D_1)}}{2D}.$$

Кривая
$$y = (e^{Dx} + e^{-Dx})/2$$
называется *цепной линией*.

4°. Мы доказали, что если решение задачи о минимальной поверхности существует, то кривая вращения — это цепная линия.

Теперь снова вернемся к обсуждению идеи Гильберта об обобщенном решении. В задаче о брахистохроне и в задаче Лопиталя решение было найдено нами. Это не обобщенное решение, а «настоящее». Правда, и в том и в другом случае экстремали не являются непрерывно дифференцируемыми функциями, если ордината одной из концевых точек равна нулю. (Впрочем, в задаче Лопиталя из такой точки свет будет распространяться «бесконечное время», так что из физических соображений точки с ординатой нуль и рассматривать не нужно.)

В задаче Дидоны было найдено решение при $l \leqslant \pi a/2$ и обобщенное решение при $l > \pi a/2$. В задаче о минимальной поверхности дело обстоит более сложным образом. Там иногда классическое решение существует, иногда — нет (см. рис. 56, где $x_0 = -a$, $x_1 = a$, $y_1 = y_0$). Если классического решения нет, то минимум доставляет обобщенное решение, состоящее из отрезков $x = -a$, $0 \leqslant y \leqslant y_0$; $x = a$, $0 \leqslant y \leqslant y_1$, соединенных отрезком $y = 0$, $-a \leqslant x \leqslant a$ (рис. 56, а). В этом случае «поверхность вращения» состоит из двух дисков, соединенных «перемычкой» $y = 0$, $-a \leqslant x \leqslant a$ (рис. 56, б). В случае же, когда экстремаль существует, надо выбрать минимум из классического и обобщенного решений.



4. Об истории вариационного исчисления. В предыдущем пункте мы объяснили содержание термина «вариационное исчисление». Теперь уместно рассказать об историческом развитии этого направления и о происхождении самого термина.

Все началось, как мы помним, с брахистохроны — задачи, поставленной Иоганном Бернулли в 1696 году. Эта задача привлекла всеобщее внимание, и вскоре было решено еще несколько подобных проблем. О некоторых упоминалось — вспомним минимальную поверхность вращения и задачу Лопиталя. И опять-таки для каждой задачи изыскивался свой способ исследования, хотя было ощущение, что возможен единый подход ко всем проблемам рассмотренного типа.

Тогда Иоганн Бернулли поставил перед своим учеником Леонардом Эйлером проблему изыскать общую методику решения всех этих задач. И Эйлер создал такой метод. В 1744 году вышел мемуар Эйлера, который так и назывался «Метод нахождения кривых линий, обладающих свойством максимума или минимума, или решение изопериметрической задачи, взятой в самом общем виде». Метод Эйлера состоял в том, что он нашел уравнение, которому должна удовлетворять «кривая линия, обладающая свойством максимума или минимума». Это уравнение уже было выписано нами. Его стали называть *уравнением Эйлера*.

Обратим внимание на слово «изопериметрический» в названии эйлеровского труда. С изопериметрической задачи мы начали нашу книгу. Какое же отношение она имеет к Эйлеру? Надо сказать, что Эйлер решил своим методом и классическую изопериметрическую задачу. В других же многочисленных случаях, которые Эйлер сумел разобрать, ограничения не имели отношения к длине кривой, к ее периметру. Но в термине «изопериметрический» отразилась преемственность нашей науки, и с тех пор он закрепился.

Знаменательное событие произошло в 1759 году. Тогда совсем еще молодой Лагранж написал свое первое сочинение, касавшееся той же темы. Он подошел к ней с другой стороны и настолько удачно, что с тех пор метод Лагранжа (называемый иногда методом вариаций) стал общепринятым. Работа привела Эйлера в восхищение, и он отказался от собственных разработок на эту тему, давая возможность молодому ученому самому довести свои замыслы до конца. Весь новый раздел математики Эйлер назвал «*вариационным исчислением*».

В чем же состоит метод Лагранжа? Что такое вариация?

Вернемся к п. 2 двенадцатого рассказа, где мы выводили конечномерную теорему Ферма. Напомним наши рассуждения.

Пусть $y = f_0(x)$ — функция n переменных $x = (x_1, \dots, x_n)$. Предположим, что она дифференцируема и достигает локального экстремума в точке $\widehat{x}$. Тогда ясно, что функция одной переменной

$$g(\lambda) = g(\lambda, j) = f_0(\widehat{x}_1, \ldots, \widehat{x}_{j-1}, \widehat{x}_j + \lambda, \widehat{x}_{j+1}, \ldots, \widehat{x}_n)$$

должна иметь минимум в нуле, и, значит, по (одномерной) теореме Ферма должно иметь место равенство

$$g'(0) = 0 \iff \frac{\partial f(\widehat{x})}{\partial x_j} = 0.$$

Лагранж в точности тот же прием применил к простейшей задаче классического вариационного исчисления:

$$\begin{gathered} F(y) = \int\limits_{x_0}^{x_1} f(x, y(x), y'(x))\,dx \to \min\ (\max), \\ y(x_0) = y_0, \quad y(x_1) = y_1. \end{gathered} \tag{1}$$

Последуем за мыслью Лагранжа. Допустим, что функция $f(x, y, z)$ в (1) непрерывно дифференцируема, а функционал $F(y)$ достигает локального минимума на непрерывно дифференцируемой кривой $\widehat{y}(x)$. Возьмем теперь «вариацию» $\widehat{y}(x)$. Точнее, возьмем любую непрерывно дифференцируемую кривую $y(x)$, которая обращается в нуль на концах: $y(x_0) = y(x_1) = 0$. Тогда «вариация» $\widehat{y}(x)$, т. е. добавка к $\widehat{y}(x)$ функции $y(x)$, умноженной на любое число λ, не выводит нас из совокупности допустимых кривых — все они будут проходить через точки (x_0, y_0) и (x_1, y_1). Следовательно, функция одной переменной

$$g(\lambda) = g(\lambda, y) = F(\widehat{y} + \lambda y) = \int\limits_{x_0}^{x_1} f(x, \widehat{y}(x) + \lambda y(x), \widehat{y}'(x) + \lambda y'(x))\,dx$$

должна иметь локальный минимум в нуле, и, значит, снова можно воспользоваться одномерной теоремой Ферма. Из этой теоремы вытекает, что $g'(0) = 0$. Теперь нужно вычислить $g'(0)$. В анализе доказывается, что при наших допущениях о непрерывной дифференцируемости $f(x, y, z)$ и $y(x)$ возможно дифференцирование под знаком интеграла. Если проделать несложные выкладки, то получается, что

$$g'(0) = \int\limits_{x_0}^{x_1} (a(x)y(x) + b(x)y'(x))\,dx, \tag{2}$$

где через $a(x)$ обозначена функция $\dfrac{\partial f}{\partial y}(x, \widehat{y}(x), \widehat{y}'(x))$, а через $b(x)$ — функция $\dfrac{\partial f}{\partial y'}(x, \widehat{y}(x), \widehat{y}'(x))$.

Итак, вслед за Лагранжем мы пришли к такому результату. Если $\widehat{y}(x)$ доставляет локальный минимум или максимум в простейшей задаче (1), то для любой непрерывно дифференцируемой функции $y(x)$, если только она обращается в нуль на концах $y(x_0) = y(x_1) = 0$, выполнено равенство:

$$\int_{x_0}^{x_1} (a(x)y(x) + b(x)y'(x))dx = 0. \tag{3}$$

Продолжим рассуждение. Найдем такую функцию $c(x)$, что $c'(x) = a(x)$ и $\int_{x_0}^{x_1} c(x)dx = \int_{x_0}^{x_1} b(x)dx = B$. Для этого подберем константу D так, чтобы интеграл по отрезку $[x_0, x_1]$ от функции $c(x) = \int_{x_0}^{x} a(\xi)d\xi + D$ равнялся бы B. А теперь проинтегрируем первое слагаемое в (3) по частям:

$$0 = \int_{x_0}^{x_1} (a(x)y(x) + b(x)y'(x))dx = \int_{x_0}^{x_1} (c'(x)y(x) + b(x)y'(x))dx = \\ = \int_{x_0}^{x_1} (b(x) - c(x))y'(x)dx.$$

Наконец, сделаем последний шаг. Положим $\bar{y}(x) = \int_{x_0}^{x} (b(\xi) - c(\xi))d\xi$. Тогда ясно, что $\bar{y}(x_0) = 0$. Кроме того, $\bar{y}(x_1) = \int_{x_0}^{x_1} (b(x) - c(x))dx = 0$ по построению функции $c(x)$, откуда $\bar{y}'(x) = b(x) - c(x)$. Это следует из формулы Ньютона—Лейбница. Значит, для $\bar{y}(x)$ должно выполняться равенство (3), т. е. $\int_{x_0}^{x_1} (b(x) - c(x))^2 dx = 0$. Но если непрерывная функция положительна, то ее интеграл не может равняться нулю. Значит, $b(x) \equiv c(x)$, т. е. доказано, что

$$b'(x) = a(x),$$

т. е.

$$\frac{d}{dx}\frac{\partial f}{\partial y'}(x, \widehat{y}(x), \widehat{y}'(x)) - \frac{\partial f}{\partial y}(x, \widehat{y}(x), \widehat{y}'(x)) \equiv 0. \tag{4}$$

Так выводится уравнение Эйлера по Лагранжу.

Вспомните: в предыдущем пункте этого рассказа говорилось, что уравнение Эйлера — это расшифровка теоремы Ферма для простейшей задачи. Здесь становится ясным смысл этого высказывания: уравнение Эйлера является следствием того, что «*производная функционала $F(y)$ в точке $\widehat{y}$ по любому направлению $y(x)$ $(y(x_0) = y(x_1) = 0)$ равна нулю*». Отметим еще, что выражение для $g'(0)$ получило название *вариации функционала F*.

Но Лагранж не остановился лишь на математических проблемах вариационного исчисления. Он, собственно говоря, и занимался-то ими лишь для того, чтобы иметь возможность применить их к проблемам естествознания. Главный труд его жизни — «Аналитическая механика» — книга о движениях тел в природе. В основе подхода Лагранжа к механике лежит экстремальный принцип, который называют *принципом наименьшего действия*. Проиллюстрируем его на простейшем примере.

Пусть шарик массы m подвешен (для простоты — в условиях невесомости) к пружине, для которой применим *закон Гука* о пропорциональности силы натяжения пружины ее отклонению от центра O. Чтобы впоследствии оставить уже введенные обозначения, мы подвесили пружину по оси Oy. Перемещение шарика тогда задает функцию $y(t)$, где $y(t)$ — его координата в момент t. Функция $y(t)$ удовлетворяет, как известно, закону Ньютона

$$my''(t) = -ky(t), \tag{5}$$

который гласит: «произведение массы на ускорение равно действующей силе» (ускорение — это вторая производная, а сила, по закону Гука, равна $-ky(t)$, где k — коэффициент пропорциональности). Выражение $T = m(y'(t))^2/2$ называется в механике *кинетической энергией*, $V = ky^2(t)/2$ — *потенциальной энергией*, а интеграл от разности кинетической и потенциальной энергий называется *действием*. Рассмотрим задачу о минимизации действия при закрепленных граничных условиях:

$$\begin{gathered}\int\limits_{t_0}^{t_1}(T-U)\,dt = \int\limits_{t_0}^{t_1}\left(\frac{my'^2}{2} - \frac{ky^2}{2}\right)dt \to \min,\\ y(t_0) = y_0, \quad y(t_1) = y_1.\end{gathered} \tag{6}$$

Если написать уравнения Эйлера для задачи (6), то мы для приходим к уравнению (5):

$$f = \frac{my'^2}{2} - \frac{ky^2}{2} \implies \frac{\partial f}{\partial y'} = my',$$

$$\frac{\partial f}{\partial y} = -ky \implies \frac{d}{dt}f_{y'} - f_y = 0 \iff my'' + ky = 0.$$

Таким образом, *второй закон Ньютона — это не что иное, как уравнение Эйлера для действия*. Или ещё: *истинные движения описываются стационарными точками действия*. При малых временах истинная траектория и на самом деле минимизирует действие. Поэтому термин «принцип *наименьшего* действия» оказывается верным при малых временах. Вообще же правильнее говорить — «принцип стационарного действия».

И снова мы столкнулись с тем, что законы природы имеют двойное описание — «физическое» и «экстремальное», а впервые разговор об этом зашел у нас в третьем рассказе и был продолжен в седьмом. Там шла речь об оптике, и минимизировалось время; здесь описываются движения тел, и при этом минимизируется действие. В оптике нами рассматривались «волновые фронты», т.е. такие совокупности точек, куда свет приходит за одно и то же время. Якоби (вслед за Гамильтоном, который с этой точки зрения исследовал оптические явления) предложил рассматривать аналоги волновых фронтов и в задачах механики (рассматривая «фронты действия»), и вообще в любой простейшей задаче классического вариационного исчисления.

А именно, Якоби начал рассматривать следующую функцию концевой точки $S(x,y)$, равной значению интеграла

$$\int_{x_0}^{x} f(\xi, y(\xi), y'(\xi))\,d\xi$$

на экстремали, доставляющей минимум, и соединяющей фиксированную точку (x_0, y_0) с точкой (x, y).

Ясно, что *любая часть экстремали, доставляющей минимум, является экстремалью, доставляющей минимум*. Это простое соображение на самом деле является расшифровкой для общей ситуации принципа Гюйгенса в оптике (о котором мы говорили в третьем рассказе), и его тоже называют *принципом Гюйгенса*. Используя принцип Гюйгенса и уравнение Эйлера, нетрудно вывести уравнение, которому удовлетворяет функция $S(x,y)$. Это уравнение называется уравнением *Гамильтона—Якоби*. Во многих интересных случаях его удается проинтегрировать, и этот метод дает еще одну возможность исследовать задачи классического вариационного исчисления. Таким образом, двойственность описания оптических явлений привела к двойному описанию решений любой задачи классического вариационного исчисления.

Вернемся несколько назад, в XVIII век. Для того чтобы можно было извлекать следствия из принципа наименьшего действия, необходимо было научиться решать задачи вариационного исчисления при ограничениях более сложных, чем изопериметрические, а именно требовалось исследовать задачи при связях, задаваемых дифференциальными уравнениями. Приведем одну общую постановку (восходящую к Лагранжу), к которой сводится большинство наиболее интересных задач из приложений. Пусть $f_j = f_j(x, y_1, \dots, y_n, z_1, \dots, z_n)$, $j = 0, 1, \dots, k$, $k < n$, функции $2n+1$ переменной. Рассмотрим задачу

$$F_0(y) = \int\limits_{x_0}^{x_1} f_0(x, y_1, \dots, y_n, y_1', \dots, y_n')dx \to \min\ (\max),$$

$$f_1(x, y_1, \dots, y_n, y_1', \dots, y_n') = 0,\ \dots,\ f_k(x, y_1, \dots, y_n, y_1', \dots, y_n') = 0$$

при граничных условиях $y_i(x_0) = y_{i0}$, $y_i(x_1) = y_{i1}$, $i = 1, \dots, n$.

Эту задачу называют *задачей Лагранжа*. Как пишутся для нее необходимые условия экстремума?

Лагранж был уверен в том, что и эта задача должна подчиняться тому принципу «снятия ограничений», который обсуждался в двенадцатом рассказе и был назван нами принципом Лагранжа. В соответствии с общим замыслом Лагранжа следует написать функцию Лагранжа и затем писать необходимое условие для задачи на экстремум функции Лагранжа при отсутствии ограничений. Но как же выглядит функция Лагранжа для задачи Лагранжа? И здесь снова Лагранж проявил максимальную решительность. В конечномерной задаче функция Лагранжа — это сумма функционала, умноженного на число λ_0, и произведений функций, задающих ограничения на множители Лагранжа. Но здесь, в задаче Лагранжа, имеется «столько же ограничений, сколько точек x на отрезке $[x_0, x_1]$», поэтому Лагранж и предложил умножать i-е уравнение $f_i(x, y_1(x), \dots, y_n(x), y_1'(x), \dots, y_n'(x)) = 0$ на функцию $l_i(x)$ аргумента x, затем интегрировать по отрезку $[x_0, x_1]$ и, наконец, суммировать по i. То есть Лагранж заменил умножение на числа и последующее суммирование умножением на функции и последующим интегрированием. В итоге функция Лагранжа приобрела следующий вид:

$$\mathscr{L} = \int\limits_{x_0}^{x_1} f(x, y_1(x), \dots, y_n(x), y_1'(x), \dots, y_n'(x))dx,$$

где

$$f = \lambda_0 f_0 + \sum_{i=1}^{k} l_i(x) f_i.$$

И он сформулировал следующий результат: если вектор-функция $\widehat{y}(x) = (\widehat{y}_1(x), \ldots, \widehat{y}_n(x))$ доставляет локальный минимум в задаче Лагранжа, то найдутся число λ_0 и функции $l_i(x)$ такие, что для функции f выполнено уравнение Эйлера. (Лагранж, конечно, умножал функционал не на λ_0, а на единицу.)

Свой результат Лагранж не доказал. Впрочем, без каких бы то ни было ограничений сформулированный результат и не может быть верным. Так что и метод, легший в основу его величайшего труда — «Аналитической механики», — остался без строгого обоснования. Это продолжалось свыше ста лет. Теорема Лагранжа была доказана с полной строгостью лишь к концу XIX века, а суть этой теоремы была осознана лишь в XX веке. А что это значит — «осознана лишь в XX веке» — это предмет особого разговора, который и составит предмет следующего пункта.

5. Заключение. (Еще раз об истории математического анализа: бесконечномерный и выпуклый анализ; теория оптимального управления и принцип максимума Понтрягина; задача о быстродействии и задача Ньютона.)

Эта книга рассчитана на школьника. В первой части я избегал элементов математического анализа. Во второй мы говорили о вещах, которые не проходят в школе, например, о функциях нескольких и бесконечного числа переменных. Но все-таки пока мы не слишком удалялись от школьной программы. И в этом, заключительном, пункте моего заключительного математического рассказа — пятнадцатый рассказ будет посвящен общим вопросам и вольным разговорам — мне тоже не хочется отходить от своей установки писать «для первого встречного» (вспомним слова Гильберта) школьника. Но здесь мне хотелось бы обрести бо́льшую свободу и мысленно обратиться лишь к такому первому встречному школьнику, который уже принял решение связать свою судьбу с математикой. Я собираюсь говорить, как бы предвидя его будущее, в расчете на то, что он впоследствии самостоятельно заполнит неизбежные пробелы понимания, вызванные недостатком его теперешних знаний.

Мне представляется важным высказать и потом частично обосновать несколько общих тезисов касательно дальнейшей судьбы идей, обсуждавшихся раньше, после чего они, с одной стороны, возносятся ввысь, а с другой — возвращаются к своим истокам.

Вспомним: сначала Ферма (для полиномов), а затем Ньютон и Лейбниц в общей ситуации разработали метод исследования задач на максимум и минимум функций одной переменной. Далее сразу началась эра классического вариационного исчисления (И. Бернулли, Эйлер, Лагранж), когда стали исследовать экстремумы некоторых функций от бесконечного числа переменных. Она продолжалась примерно два с половиной века.

В конце XIX века Вольтерра и чуть позже — Фреше, Адамар и многие другие стали развивать основания бесконечномерного анализа. При этом всегда подчеркивалось, что одна из целей вновь создаваемого исчисления — решать бесконечномерные задачи на максимум и минимум. Вспомним слова Вольтерра, поставленные эпиграфом к нашему рассказу. В первой половине нашего столетия математический анализ в бесконечномерных пространствах (его стали называть *функциональным анализом*, этот раздел анализа соединил в себе различные концепции классического анализа, высшей алгебры и геометрии) пережил пору бурного развития и расцвета. Однако математики, которые в это же время продолжали развивать вариационное исчисление, не прилагали общих теорем функционального анализа к вариационному исчислению и, более того, не считали, что разрабатывается аппарат для этой теории. Один из крупнейших специалистов в области вариационного исчисления в XX века — Блисс — в своем учебнике по вариационному исчислению, где он подвел итоги всего развития этой дисциплины, в книге, создававшейся в сороковые годы, когда все необходимые результаты функционального анализа уже стали общеизвестными, очень скептически отзывался о возможностях этого общего подхода в «конкретных» задачах вариационного исчисления.

И вот, пожалуй, теперь настало время сформулировать наш первый тезис.

Бесконечномерный анализ (точнее, дифференциальное исчисление в бесконечномерных пространствах) — *это раздел математики, основанный в точности на тех же самых идеях, что и конечномерный, и столь же простой и естественный, как последний. Он является истинным аппаратом классического вариационного исчисления в той же мере, в какой конечномерный является аппаратом для теории конечномерных задач на экстремум*. При этом все основные теоремы дифференциального исчисления в бесконечномерном пространстве столь же просты и естественны, как их конечномерные аналоги.

В основании дифференциального и интегрального исчисления для функций одной переменной лежат понятия *производной* и *дифференциала*. Вспомним их определение (рассказ одиннадцатый).

Мы говорили, что функция $F(x)$, заданная на прямой $\mathbb{R}$, *дифференцируема в точке* x_0, если существует такая линейная функция $y = kx$, что

$$F(x_0 + x) - F(x_0) = kx + r(x),$$

где $\lim\limits_{|x|\to 0} |r(x)|/|x| = 0$. Линейная функция $y = kx$ называется *дифференциалом* F *в точке* x_0. Бесконечномерный анализ изучает функции на *пространствах с нормой* — так называемых *нормированных пространствах*. Примером нормированных пространств могут служить встретившиеся нам пространства $C[a,b]$ и $C^1[a,b]$. А вообще нормированным пространством называется любая совокупность Y таких элементов, с которыми можно поступать так же, как с векторами на плоскости, т. е. складывать их и умножать на числа, а кроме того, таких, что каждому элементу y из Y приписано число $\|y\|$, называемое *нормой*. Норма должна удовлетворять таким свойствам:

1) $\|y\| \geqslant 0$ (неотрицательность) и $\|y\| = 0$, лишь если $y = 0$;

2) $\|ay\| = |a| \cdot \|y\|$ для любых $y \in Y$ и $a \in \mathbb{R}$ (однородность);

3) $\|y_1 + y_2\| \leqslant \|y_1\| + \|y_2\|$ для всех y_1, y_2 из Y (аксиома треугольника).

Примерами нормированных пространств могут служить прямая $\mathbb{R}$, где следует положить $\|y\| = |y|$ n-мерное пространство векторов $y = (y_1, \ldots, y_n)$, где можно ввести нормы по-разному, например,

$$\|y\| = |y_1| + \ldots + |y_n|, \quad \|y\| = \sqrt{y_1^2 + \ldots + y_n^2} \quad \text{и т. п.},$$

а также встретившиеся нам пространства C и C^1.

Мы уже говорили, что функция $K(y)$ называется *линейной*, если $K(y_1+y_2) = K(y_1)+K(y_2)$ для всех $y_1, y_2 \in Y$ и $K(ay) = aK(y)$ для любых $y \in Y$, $a \in \mathbb{R}$.

Теперь можно определить понятие производной в любом случае.

Мы говорим, что функция $F(y)$, заданная на нормированном пространстве Y, *дифференцируема в точке* y_0, если существует такая линейная функция $K(y)$, что

$$F(y_0 + y) - F(y_0) = K(y) + r(y),$$

где $\lim\limits_{\|y\| \to 0} |r(y)|/\|y\| = 0$.

Линейная функция $K(y)$ называется *дифференциалом F в точке y_0*.

Не правда ли, все очень похоже? Скажу больше. Определение дифференциала в бесконечномерном анализе дал французский математик Фреше в начале XX века. Посмотрим, к чему приведет определение Фреше, если применить его к конечномерному случаю.

Мы скажем, что функция $F(y) = F(y_1, \ldots, y_n)$ n переменных *дифференцируема в точке* $y_0 = (y_{01}, \ldots, y_{0n})$, если существует такая линейная функция (а любая такая функция имеет, как мы помним, вид $K(y) = k_1 y_1 + \ldots + k_n y_n$), что

$$F(y_{01} + y_1, \ldots, y_{0n} + y_n) - F(y_{01}, \ldots, y_{0n}) = K(y) + r(y),$$

где $\lim\limits_{\|y\| \to 0} |r(y)|/\|y\| = 0$. В качестве нормы можно взять любую норму в n-мерном пространстве. Сейчас именно это определение входит в каждый учебник по анализу. Но Фреше считал, что его придумал именно он! Он так и пишет: «дифференциал в моем смысле». Обычный конечномерный анализ существовал уже два с половиной столетия, и вместе с тем один из крупнейших математиков начала XX века считает, что он впервые предлагает правильное определение основного понятия анализа — понятия дифференциала! Правда, потом выяснилось, что «его» определение уже было дано в неопубликованных трудах Вейерштрасса (относящихся к 60-м годам XIX века) и встречалось в английских и немецких учебниках начала века, но в научной литературе его действительно не было. Факт остается фактом — бесконечномерное определение оказалось легче придумать, чем конечномерное!

Далее. Какие теоремы дифференциального исчисления следует считать основными? Прежде всего необходимо назвать две теоремы — о дифференцировании сложной функции и об обратной функции.

Правило дифференцирования сложной функции (одной переменной) таково:

$$(F(G(x)))' = F'(G(x))G'(x).$$

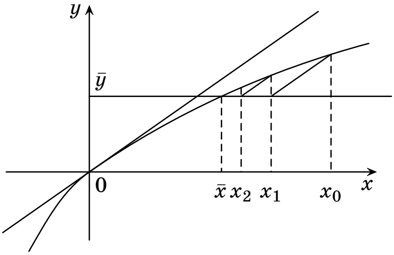
В бесконечномерном анализе эта формула также верна (если придать ей правильную трактовку) и доказывается столь же просто, как и в конечномерном.

Теорема об обратной функции (для одной переменной) гласит: если непрерывно дифференцируемая функция $y = f(x)$ равна нулю в точке 0 и при этом $f'(0) \neq 0$, то вблизи нуля функцию f можно обратить, т. е. для любого достаточно малого $\bar{y}$ найдется такое единственное $\bar{x}$, что $\bar{y} = f(\bar{x})$.

Как доказывается эта теорема? Среди разнообразных доказательств ее отдадим предпочтение доказательству, восходящему к Ньютону. Оно состоит в построении последовательности $\{x_n\}_{n\geqslant 0}$, постепенно сходящейся к $\bar{x}$. Эта последовательность строится по правилу:

$$x_{k+1} = x_k + (f'(0))^{-1}(\bar{y} - f(x_k)), \quad k \geqslant 0, \tag{1}$$

а геометрически она изображена на рис. 57. Нулевое приближение x_0 берется произвольно в достаточной близости от нуля.



В бесконечномерном случае верен тот же результат. Рискну сформулировать его. Одной нормированности здесь недостаточно. Нужно еще одно свойство, так называемое свойство полноты, «отсутствия пробелов» (рациональные числа не образуют полного пространства, ибо между ними имеются «пробелы» типа числа $\sqrt{2}$, которое не рационально). Полные нормированные пространства называются *банаховыми*.

Пусть теперь X и Y — два банаховых пространства и F — функция, отображающая X в Y. Производная ее — это линейное непрерывное отображение, переводящее X в Y. Условие $f'(0) \neq 0$ в одномерной теореме на наш «бесконечномерный язык» переводится так: производная $F'(0)$ обратима (т. е. $(F'(0))^{-1}$ является также линейным и непрерывным отображением из Y в X). И теперь теорема об обратной функции получает совершенно аналогичный вид: если X и Y — банаховы пространства, непрерывно дифференцируемая функция $y = F(x)$ равна нулю в нуле и $F'(0)$ — обратима, то для любого $\bar{y}$, близкого к нулю, найдется такое единственное $\bar{x}$, что $F(\bar{x}) = \bar{y}$.

И доказательство фактически никак не меняется: доказательство сходимости x_k, определенных по формуле (1), к $\bar{x}$ проводится совершенно аналогично тому, как на прямой $\mathbb{R}$.

Подведем итоги. И основное понятие — понятие дифференциала, и формулировки основных теорем, и их смысл, и их доказательства в одномерном и бесконечномерном случае фактически одинаковы. При этом можно сказать следующее.

В бесконечномерном анализе верен принцип Лагранжа, причем основания для него те же, что в конечномерном случае, и доказательства столь же просты.

Пусть снова X и Y — банаховы пространства, f_i — функционалы на X, $i = 0, 1, \ldots, m$, а F — отображение из X в Y. Рассмотрим задачу

$$(\text{з}) \qquad f_0(x) \to \min(\max),$$
$$F(x) = 0, \quad f_i(x) = 0, \quad i = 1, \ldots, m.$$

Если отображение F и функции f_i отсутствуют, то получается задача без ограничений:

$$(\text{з}') \qquad f_0(x) \to \min\ (\max).$$

В применении к задаче (з) принцип Лагранжа формулируется с употреблением тех же слов, что употребляются и в конечномерном случае, и для изопериметрических задач классического вариационного исчисления, а именно:

1) Для задачи (з$'$) необходимым условием экстремальности является соотношение

$$f_0'(\widehat{x}) = 0$$

(теорема Ферма).

2) Для решения задачи (з) надо *составить функцию Лагранжа задачи и затем поступить с ней так, как будто переменные независимы.*

Функция же Лагранжа задачи (з) имеет стандартный вид

$$\mathscr{L} = \mathscr{L}(x, \lambda_0, \ldots, \lambda_m, \Lambda) = \lambda_0 f_0(x) + \ldots + \lambda_m f_m(x) + \Lambda(F(x)),$$

где $\lambda_0, \ldots, \lambda_m$ — числа, а $\Lambda(y)$ — линейная функция на Y.

Таким образом, если $\widehat{x}$ — локальный экстремум, то найдутся такие числа $\lambda_0, \ldots, \lambda_m$ и такая линейная функция $\Lambda(y)$, что выполнена теорема Ферма для функции Лагранжа:

$$\frac{\partial \mathscr{L}}{\partial x}(\widehat{x}, \lambda_0, \ldots, \lambda_m, \Lambda) = 0.$$

Но именно так и формулируется правило множителей Лагранжа в конечномерном случае — вспомним соответствующую теорему из двенадцатого рассказа. Сформулированную теорему о правиле множителей Лагранжа доказал в 1934 году советский математик Л. А. Люстерник.

При этом необходимо сказать, что в бесконечномерном случае, помимо гладкости входящих в постановку задачи функций и отображений (а ведь это было единственным требованием в конечномерном случае), нужны еще два требования.

Во-первых, нужно, чтобы пространства X и Y были полными, т. е. банаховыми, во-вторых, нужно, чтобы отображение F обладало бы некоторыми особыми свойствами, которые обычно называют свойствами регулярности (для регулярности и, следовательно, для того, чтобы было справедливо правило множителей Лагранжа, достаточно, чтобы производная $F'(\widehat{x})$ отображала X на все Y). Доказывается бесконечномерный результат столь же просто, как и конечномерный. Читателя, который пожелает в этом убедиться, я отсылаю к книге [2], где доказательство конечномерного случая приведено на с. 49—50, а бесконечномерного — на с. 254—255. Оба доказательства строго параллельны, только в нескольких местах некоторые известные результаты высшей алгебры и классического анализа заменяются их обобщениями в функциональном анализе (эти обобщения входят в самый минимум математического университетского образования).

А что же такое *выпуклость и выпуклый анализ*? С понятием выпуклости мы знакомимся в школе на уроках геометрии. Напомню: *фигура* на плоскости или в пространстве *называется выпуклой*, если вместе с двумя своими точками она содержит весь отрезок, соединяющий эти точки; *функция называется выпуклой*, если ее график лежит не ниже хорды, соединяющей любые две точки графика. На рис. 45 изображены выпуклые и невыпуклые фигуры и функции. Треугольник — всегда выпуклое множество, а среди четырехугольников встречаются и невыпуклые. Линейные ($y = ax$, $y = a_1x_1 + \ldots + a_nx_n$) и аффинные ($y = ax+b$, $y = a_1x_1+\ldots+a_nx_n+b$) функции являются выпуклыми;

среди квадратных трехчленов $y = ax^2 + bx + c$, $a \neq 0$, выпуклыми являются лишь те, у которых $a > 0$. Выпуклые функции можно определить аналитически. Функция $y = f(x)$ является выпуклой в том и только в том случае, если для любых точек x_1 и x_2 и любого числа α, лежащего между нулем и единицей, имеет место *неравенство Иенсена*

$$f(\alpha x_1 + (1-\alpha)x_2) \leqslant \alpha f(x_1) + (1-\alpha)f(x_2).$$

Это неравенство обобщается на случай n точек:

$$f(\alpha_1 x_1 + \ldots + \alpha_n x_n) \leqslant \alpha_1 f(x_1) + \ldots + \alpha_n f(x_n)$$

(если только $\alpha_1 \geqslant 0, \ldots, \alpha_n \geqslant 0$, $\sum\limits_{i=1}^{n} \alpha_i = 1$).

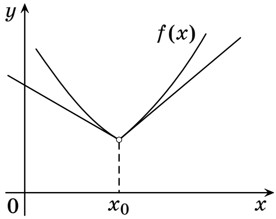
Вот важное достаточное условие выпуклости функции одной переменной: если функция f дважды дифференцируема и при этом $f''(x) \geqslant 0$ для всех x, то она выпукла. Посмотрим теперь на функции из таблицы, приведенные в одиннадцатом рассказе. Легко понять, что функция $y = |x|^a$ выпукла тогда и только тогда, когда $a \geqslant 1$, функция $y = a^x$ $(a > 0,\ a \neq 1)$ всегда выпукла, выпуклыми являются и функции $y = \log_a x$, $0 < a < 1$, $y = -\ln x$. Функции $y = \sin x$ и $y = \cos x$ выпуклыми не являются.

Среди всех фигур выпуклые составляют лишь достаточно узкий, специальный класс. То же можно сказать и о выпуклых функциях. Однако выпуклость играет очень важную роль и в математике, и в ее приложениях. Содержательность идей, сопутствующих выпуклости, и богатство приложений привели к созданию отдельного раздела в математике, получившего название «выпуклый анализ». Его окончательное оформление произошло сравнительно недавно — лет сорок тому назад. В этом разделе изучают свойства выпуклых множеств, выпуклых функций и выпуклых экстремальных задач. Нас в основном будут интересовать, конечно, именно экстремальные задачи.

Надо сказать, что именно обилие выпуклых экстремальных задач привело к необходимости углубленного изучения выпуклости и, в итоге, — к созданию выпуклого анализа. Особенно много выпуклых задач возникает в экономике. Об одной задаче экономического содержания — транспортной задаче — мы говорили. Такого рода проблемы стали возникать постоянно. Очень часто при формализации таких задач выяснялось, что функции, которые требовалось максимизировать или минимизировать, а также функции, задающие ограничения (типа равенств и неравенств), являются линейными. Возник вопрос о том, как их решать. Методы их решения составили специальный раздел выпуклого анализа — линейное программирование. Первые работы, относящиеся к этому разделу, принадлежат советскому математику — академику, лауреату Ленинской и Нобелевской премий Л. В. Канторовичу.

Что же изучают в выпуклом анализе? Важный раздел выпуклого анализа составляет так называемое «выпуклое исчисление», имеющее много сходного с дифференциальным исчислением. Объясним, что это такое. Не все выпуклые функции дифференцируемы. Примером выпуклой функции, не дифференцируемой в нуле, является функция $y = |x|$.

Мы уже объясняли, что у этой функции нет (в нуле) касательной. Но у любой выпуклой функции $y = f(x)$ одной переменной имеются две «полукасательные» (рис. 58).



Это означает, что всегда существуют пределы (далее $h > 0$)

$$f'_-(x_0) = \lim_{h>0,\, h\to 0} \frac{f_0(x_0 - h) - f(x_0)}{-h},$$

$$f'_+(x_0) = \lim_{h>0,\, h\to 0} \frac{f_0(x_0 + h) - f(x_0)}{h}.$$

Весь отрезок $[f'_-(x_0), f'_+(x_0)]$ (как правило, вырождающийся в точку) называется *субдифференциалом* функции f в точке x_0. Он обозначается $\partial f(x_0)$. Если f — константа, то $\partial f(x) = 0$, если $f(x) = ax+b$, то $\partial f(x) = a$ в любой точке x. И вообще, если $y = f(x)$ — функция, дифференцируемая в точке x, то $\partial f(x_0) = f'(x_0)$.

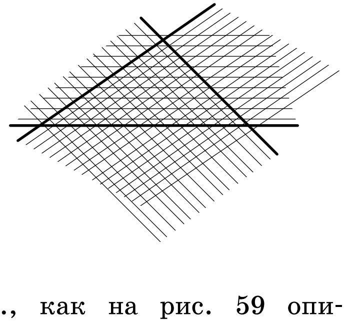
Если дифференцируемая функция имеет локальный минимум в точке x_0, то, как мы знаем, по теореме Ферма имеет место равенство $f'(x_0) = 0$. Мы отмечали при этом, что это равенство является необходимым, но не достаточным условием экстремума. Если функция выпукла, то ее локальный минимум всегда глобальный, абсолютный. В этом состоит одно из замечательных свойств выпуклых функций. Необходимым и достаточным условием минимальности выпуклой функции f в точке x_0 является соотношение $0 \in \partial f(x_0)$ (теорема Ферма для выпуклых функций). Это соотношение означает совсем простую вещь: для того чтобы выпуклая функция f достигала минимума в точке x_0, необходимо и достаточно, чтобы постоянная функция, равная $f(x_0)$, лежала бы ниже графика f.

Понятие субдифференциала можно обобщить и на случай функции n переменных. Субдифференциал, в отличие от производной, — это не вектор, а некоторое выпуклое множество векторов. При этом имеют место формулы, похожие на формулы дифференциального исчисления. Например, имеет место такая формула:

$$\partial(f+g)(x) = \partial f(x) + \partial g(x).$$

Она обобщает формулу $(f+g)'(x) = f'(x) + g'(x)$, о которой говорилось в одиннадцатом рассказе. Приведенная формула для субдифференциала суммы означает, что для того, чтобы узнать субдифференциал функции $f+g$ в точке x, надо взять множества $A = \partial f(x)$, $B = \partial g(x)$ и образовать множество $A+B$ из суммы $a+b$, $a \in A$, $b \in B$.

Соотношения, подобные формуле для субдифференциала суммы, и образуют *выпуклое исчисление*.



Важнейший тезис выпуклого анализа состоит в том, что выпуклые объекты всегда допускают двойное описание и что для выпуклого объекта всегда имеется «двойственный» объект. Скажем, выпуклая фигура на плоскости может быть описана сама по себе, как совокупность своих точек, а может быть описана также и как пересечение всех содержащих полуплоскостей (полупространств) (см., как на рис. 59 описывается треугольник). Аналогично, всякая выпуклая функция может быть описана сама по себе, а также как максимум из всех аффинных функций, ее не превосходящих.

Это последнее двойственное описание приводит к одному из важнейших понятий всего классического анализа — понятию *преобразования Лежандра*.

Пусть $y = f(x)$ — выпуклая функция. Ее преобразованием

Лежандра называется функция

$$y = f^*(z), \quad f^*(z) = \max_x (xz - f(x)).$$

Приведем пример. Пусть $f_p(x) = |x|^p/p$, $p > 1$. (При $p > 1$ функция $y = f_p(x)$ является, как мы помним, выпуклой.) Для того чтобы найти $\max(xz - f_p(x))$, применим теорему Ферма.

Имеем $(d/dx)(xz - f_p(x)) = z - |x|^{p-1}\operatorname{sign} x = 0$. Тогда $x = |z|^{p'-1} \times \operatorname{sign} z$, где $(p')^{-1} + p^{-1} = 1$. При этом

$$f_p^*(z) = zx - |x|^p/p = |z|^{p'} - (x|x|^{p-1}\operatorname{sign} x)/p = \\ = |z|^{p'}(1 - 1/p) = |z|^{p'}/p' =: f_{p'}(z).$$

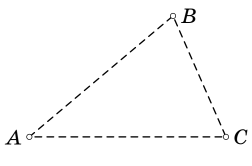
Таким образом, функции $f_p(x)$ и $f_{p'}(x)$ двойственны: каждая из них является преобразованием Лежандра другой. При этом функция $f_2(x) = x^2/2$ является «самодвойственной».

В этом обстоятельстве лежит причина неравенств Коши—Буняковского и Гельдера, о которых говорилось в рассказе пятом. Отметим, кстати, что большинство неравенств, исследовавшихся там, имеет отношение к выпуклому анализу. Например, неравенство между средним арифметическим и средним геометрическим есть не что иное, как неравенство Иенсена. Действительно, пусть $x_i > 0$, $i = 1, \ldots, n$. Тогда

$$(x_1 \ldots x_n)^{1/n} = e^{\frac{\ln x_1 + \ldots + \ln x_n}{n}} \leqslant \frac{1}{n}(e^{\ln x_1} + \ldots + e^{\ln x_n}) = \frac{x_1 + \ldots + x_n}{n}.$$

Мы воспользовались неравенством Иенсена для выпуклой функции $y = e^x$.

И еще один важнейший тезис выпуклого анализа. Рассмотрим треугольник ABC (см. рис. 60). Мы можем стереть все, кроме его вершин. Ведь зная вершины, мы восстановим весь треугольник. То же можно сказать относительно квадрата, ромба и вообще любого выпуклого многоугольника — они восстанавливаются по своим вершинам. Вершины многоугольника характеризуются тем, что они (в отличие от остальных точек многоугольника) не являются серединами каких-либо отрезков с концами в точках этого многоугольника.



Оказывается, у любого ограниченного (и замкнутого) выпуклого множества существуют «крайние» точки, т. е. такие точки, которые не могут оказаться серединами никаких отрезков, принадлежащих этому множеству. У круга, например, множеством крайних точек является ограничивающая его окружность. И при этом оказывается, что всякое выпуклое множество восстанавливается по своим крайним точкам.

Напомним, что задачи линейного программирования формализуются так:

$$a_{01}x_1 + \ldots + a_{0n}x_n \to \min \ (\max),$$
$$a_{i1}x_1 + \ldots + a_{in}x_n \leqslant b, \quad i = 1, \ldots, m,$$
$$x_i \geqslant 0, \quad i = 1, \ldots, n.$$

Ограничения в таких задачах образуют «многогранник», который восстанавливается по множеству своих вершин. При этом, как легко понять, минимум (или максимум) линейной функции достигается в одной из вершин.

Один из наиболее широко известных в мире численных методов (так называемый *симплекс-метод*) устроен так, что он дает возможность из некоторой вершины переходить в другую с меньшим (бóльшим) значением минимизируемой (максимизируемой) функции. Через конечное число шагов находится нужный экстремум.

Теорию выпуклых экстремальных задач называют *выпуклым программированием*. Достаточно широкий класс конечномерных задач выпуклого программирования описывает следующая постановка:

$$f_0(x) \to \min, \qquad (з')$$
$$f_i(x) = 0, \quad i = 1, \ldots, m',$$
$$f_i(x) \leqslant 0, \quad i = m'+1, \ldots, m, \quad x \in A.$$

По сравнению с той постановкой, которую мы привели в начале параграфа, здесь имеются следующие изменения. Во-первых, это задача на минимум, во-вторых, функция f_0 и функции, задающие неравенства, должны быть выпуклыми, в третьих, функции, задающие равенства, должны быть аффинными и, наконец, появилось еще одно ограничение: $x \in A$. При этом множество A также предполагается выпуклым. Вот в этом случае — для выпуклых задач — слова Лагранжа, сказанные о своем принципе и помещенные нами в эпиграф, полностью оправдываются, а именно, имеет место следующая теорема.

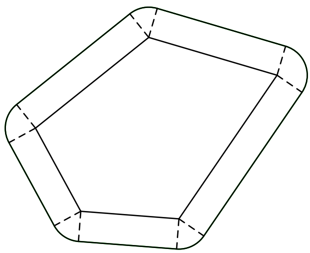
Если допустимая точка $\widehat{x}$ доставляет абсолютный минимум в задаче (з′) *выпуклого программирования, то найдутся* *числа* $\lambda_0, \ldots, \lambda_m$, *не равные одновременно нулю и такие, что* а) *выполнены условия неотрицательности* $\lambda_0 \geqslant 0$, $\lambda_i \geqslant 0$, $i \geqslant m' + 1$; б) *выполнены условия дополняющей нежесткости* $\lambda_i f_i(\widehat{x}) = 0$, $i \geqslant m' + 1$; *и, наконец,* в) *имеет место так называемый принцип минимума, согласно которому точка* $\widehat{x}$ *является точкой минимума функции Лагранжа в задаче*

$$\mathscr{L}(x, \lambda_0, \ldots, \lambda_m) \to \min, \quad x \in A, \quad \mathscr{L} = \lambda_0 f_0(x) + \ldots + \lambda_m f_m(x).$$

Если же оказалось, что нашлись такие числа $\lambda_1, \ldots, \lambda_m$ *и* $\lambda_0 = 1$, *что удовлетворяются соотношения* а)—в), *то* $\widehat{x}$ *является абсолютным минимумом в задаче.*

Эта теорема была доказана сравнительно недавно — в 1951 году — американскими математиками Куном и Таккером. Она играет роль правила множителей Лагранжа для задач выпуклого программирования.

Закончим наш рассказ решением задачи Дидоны средствами выпуклого анализа.



Пусть A_1 и A_2 — две выпуклые фигуры на плоскости. Через $\alpha_1 A_1 + \alpha_2 A_2$ обозначим фигуру, образованную векторами x, представимыми в виде $x = \alpha_1 x_1 + \alpha_2 x_2$, где x_1 из A_1, x_2 из A_2, $\alpha_1 \geqslant 0$, $\alpha_2 \geqslant 0$. Сумма многоугольника с кругом изображена на рис. 61.

Обозначим через $S(A)$ площадь множества A. В середине прошлого века немецкий геометр Брунн доказал следующее важное неравенство:

$$\sqrt{S(\alpha A_1 + (1-\alpha)A_2)} \geqslant \alpha\sqrt{S(A_1)} + (1-\alpha)\sqrt{S(A_2)}; \tag{1}$$

затем в конце XIX века Минковский доказал, что равенство в (1) возможно тогда и только тогда, когда A_1 и A_2 гомотетичны. Доказательство этих фактов и их обобщений, известных под названием неравенств Брунна—Минковского, можно найти в упоминавшейся нами книге Люстерника [14].

Пусть теперь A — выпуклая фигура, а B — единичный круг. Тогда имеет место такая интересная формула (доказанная Штейнером):

$$S(A + \rho B) = S(A) + p(A)\rho + \pi\rho^2. \tag{2}$$

В (2) $p(A)$ — периметр (длина границы) A. Для выпуклого многоугольника формула (2) очевидна (см. рис. 61). В общем случае ее можно доказать предельным переходом.

Покажем, как из формул (1) и (2) сразу следует решение классической изопериметрической задачи. Далее мы пользуемся тем, что $p(\alpha A) = \alpha p(A)$ и $S(\alpha A) = \alpha^2 S(A)$. Имеем

$$\sqrt{S(\alpha A + (1-\alpha)B)} \overset{(2)}{=} \sqrt{S(\alpha A) + p(\alpha A)(1-\alpha) + \pi(1-\alpha)^2} =$$
$$= \sqrt{\alpha^2 S(A) + \alpha(1-\alpha)p(A) + \pi(1-\alpha)^2} \overset{(1)}{\geqslant}$$
$$\overset{(1)}{\geqslant} \alpha\sqrt{S(A)} + (1-\alpha)\sqrt{S(B)} = \alpha\sqrt{S(A)} + (1-\alpha)\sqrt{\pi}.$$

Возводя в квадрат и сокращая одинаковые члены, приходим к изопериметрическому неравенству $p^2(A) \geqslant 4\pi S(A)$, которое было получено в итоге второго рассказа. При этом равенство достигается, лишь если A — круг. Вот и еще одно решение задачи Дидоны.

Вернемся снова в XVIII век.

Мы как-то сказали, что Лагранж выдвинул, активно применял и пропагандировал свой принцип снятия ограничений для вариационных задач, но не доказал его. Обоснование этого принципа Лагранжа для задач классического вариационного исчисления стало уделом математиков более поздних поколений, оно было дано в конце XIX века (Майер и др.). И вместе с тем, теорема Лагранжа прямо следует из основных теорем бесконечномерного анализа.

Как же вписывается классическое вариационное исчисление в бесконечномерный анализ? Приведем один частный, но исключительный по важности случай общей задачи Лагранжа.

$$I_0(y_1, \dots, y_n, u_1, \dots, u_r) =$$
$$= \int_{x_0}^{x_1} f_0(x, y_1(x), \dots, y_n(x), u_1(x), \dots, u_r(x))\,dx \to \min\ (\max),$$
$$y_1' = \varphi_1(x, y_1, \dots, y_n, u_1, \dots, u_r),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$y_n' = \varphi_n(x, y_1, \dots, y_n, u_1, \dots, u_r)$$

при граничных условиях $y_i(x_0) = y_{i0},\ y_i(x_1) = y_{i1},\ i = 1, \dots, n.$

В этой постановке неизвестные функции разделены на два класса. Одни $(y_1(x), \dots, y_n(x))$ связаны дифференциальными уравнениями. Другие $(u_1(x), \dots, u_r(x))$ можно выбирать свободно. Их называют *управлениями*. Схема дифференциального уравнения или системы таких уравнений с управляющими функциями описывает множество явлений, где допустимо вмешательство человека.

Если считать (при условии, что $f_0, \varphi_1, \ldots, \varphi_n$ непрерывно дифференцируемы) все игреки непрерывно дифференцируемыми, а управления $u_1(x), \ldots, u_r(x)$ непрерывными, можно рассмотреть то отображение, которое сопоставляет набору функций $y_1(x), \ldots$ $\ldots, y_n(x), u_1(x), \ldots, u_r(x)$ набор непрерывных функций

$$\begin{gathered} z_1(x) = y_1'(x) - \varphi_1(x, y_1(x), \ldots, y_n(x), u_1(x), \ldots, u_r(x)), \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ z_n(x) = y_n'(x) - \varphi_n(x, y_1(x), \ldots, y_n(x), u_1(x), \ldots, u_r(x)). \end{gathered}$$

Запишем его как $z(x) = F(y_1(x), \ldots, u_r(x))$, $z(x) = (z_1(x), \ldots$ $\ldots, z_n(x))$. При этом сами дифференциальные уравнения в (з$_1$) могут рассматриваться как равенство $F(y_1(x), \ldots, u_r(x)) = 0$. В итоге получается задача вида

$$I_0(y_1, \ldots, u_r) \to \min\ (\max), \quad F(y_1(x), \ldots, u_r(x)) \equiv 0$$

с краевыми условиями $y_i(x_0) = y_{i0}$, $y_i(x_1) = y_{i1}$, $i = 1, \ldots, n$.

И к этой задаче оказывается возможным применить мысль Лагранжа. Функция Лагранжа здесь принимает вид

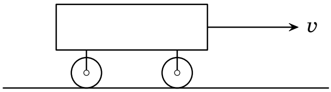
$$\begin{aligned} \mathscr{L} = \lambda_0 I + \int_{x_0}^{x_1} (p_1(x)(y_1' - \varphi_1(x, y_1, \ldots, u_r)) + \ldots \\ \ldots + p_n(x)(y_n' - \varphi_n(x, y_1, \ldots, u_r)))\,dx. \end{aligned}$$

Далее следует рассмотреть задачу

$$\begin{gathered} \mathscr{L} \to \min\ (\max), \\ y_i(x_0) = y_{i0}, \quad y_i(x_1) = y_{i1}, \quad i = 1, \ldots, n, \end{gathered}$$

и поступать с нею так, как будто ограничений нет, т. е. писать систему уравнений Эйлера. А справедливость принципа Лагранжа в этом случае выводится из теоремы Люстерника, о которой упоминалось выше.

В заключение расскажем о задачах оптимального управления. Этот важнейший раздел теории экстремальных задач был вызван к жизни задачами технического содержания. Вспомним простейшую задачу оптимального управления из первого рассказа. Пусть имеется тележка, движущаяся прямолинейно без трения по горизонтальным рельсам (см. рис. 62). Тележка управляется внешней силой, которую можно изменять в заданных пределах. Требуется остановить тележку в определенном положении в заданное время. Эту задачу мы называем далее простейшей задачей о быстродействии. Формализуем ее.



Пусть масса тележки m, ее начальная координата x_0, а начальная скорость v_0. Внешнюю систему (силу тяги) обозначим через u, а текущую координату тележки — через $x(t)$. Тогда скорость тележки $v(t) = dx(t)/dt$, а ускорение $a(t) = d^2x(t)/dt^2$. Конечный момент времени обозначим через T. По закону Ньютона $m \cdot a(t)$ есть внешняя сила, которая равна $u(t)$. Ограничение на внешнюю силу зададим в виде неравенств $u_1 \leqslant u(t) \leqslant u_2$. В итоге получается такая формализация:

$$T \to \min, \quad \frac{dx}{dt} = v, \quad m\frac{dv}{dt} = u, \quad x(0) = x_0, \quad \frac{dx(0)}{dt} = v_0,$$
$$x(T) = \frac{dx(T)}{dt} = 0, \quad u_1 \leqslant u \leqslant u_2.$$

Если бы не было ограничения типа неравенств, эта задача относилась бы к классическому вариационному исчислению. Но наличие ограничения типа нестрогих неравенств на управления не позволяет применять разработанные в вариационном исчислении методы.

Задачи с такими ограничениями получили название задач *оптимального управления*. Их теория была разработана советским ученым Понтрягиным и его учениками Болтянским, Гамкрелидзе, Мищенко.

Основной метод решения таких задач был открыт в их работах. Он получил название *принципа максимума Понтрягина*.

Что это такое? Сама задача оптимального управления формулируется почти так же, как (з$_1$), только на управления налагаются еще дополнительные ограничения, которые мы запишем так:

$$(u_1(x), \ldots, u_r(x)) \in U, \qquad (*)$$

где U — это какое-то фиксированное множество векторов $(u_1, \ldots$ $\ldots, u_r)$.

Вот самый простой пример задачи оптимального управления.

Пусть $f(x,u)$ — непрерывная функция двух переменных. Рассмотрим задачу

$$(\text{з}_2) \quad \int\limits_{x_0}^{x_1} f(x,u(x))\,dx \to \min, \quad u(x) \in [u_1,u_2] \ (\Longleftrightarrow u_1 \leqslant u(x) \leqslant u_2).$$

Ограничимся частным случаем такой задачи:

$$\int\limits_{x_0}^{x_1} p(x)u(x)\,dx \to \min, \quad -1 \leqslant u(x) \leqslant 1.$$

Здесь $p(x)$ — некоторая непрерывная функция на отрезке $[x_0,x_1]$.

Как нам поступить? Совсем нетрудно понять, что наш интеграл будет самым маленьким, если в случае $p(x) > 0$ мы положим $\widehat{u}(x) = -1$, а в случае $p(x) < 0$ положим $u(x)$ равным $+1$, т. е. $\widehat{u}(x) = -\operatorname{sign} p(x)$. И в общей задаче (з$_2$) следует поступить аналогичным образом, а именно, надо для каждого x из $[x_0,x_1]$ найти то u из отрезка $[u_1,u_2]$, при котором функция $f(x,u)$ (по u) имеет на этом отрезке минимум.

Это можно сформулировать в виде следующего принципа минимума: для того чтобы функция $\widehat{u}(\cdot)$ была решением задачи (з$_2$), необходимо, чтобы

$$\min_{u\in[u_1,u_2]} f(x,u) = f(x,\widehat{u}(x)).$$

А теперь можно объяснить, в чем состоит принцип максимума. К задаче оптимального управления также применим общий замысел Лагранжа, только его надо чуть модифицировать. Если нам нужно решить задачу оптимального управления (з$_2$) с дополнительным ограничением $(*)$, то нужно составить функцию Лагранжа $\mathscr{L}$ (не отразив в ней ограничение $(*)$), затем рассмотреть задачу

$$\mathscr{L} \to \min, \quad y_i(x_0) = y_{i0}, \quad y_i(x_1) = y_{i1},$$
$$i = 1,\dots,n, \quad (u_1,\dots,u_m) \in U.$$

При этом по y надо поступить как и раньше, т. е. составить уравнения Эйлера, а по u применить принцип минимума. Но так как все члены с u в функцию Лагранжа вошли со знаком «$-$», удобнее записать его в виде *принципа максимума*, где $(\widehat{y}_1(x), \ldots, \widehat{y}_n(x), \widehat{u}_1(x), \ldots, \widehat{u}_r(x))$ — решение задачи

$$\max_{(u_1,\ldots,u_r)\in U} (p_1(x)\varphi_1(x, \widehat{y}_1(x), \ldots, \widehat{y}_n(x), u_1, \ldots u_r) + \ldots$$
$$\ldots + p_n(x)\varphi_n(x, \widehat{y}_1(x), \ldots, \widehat{y}_n(x), u_1, \ldots, u_r)) =$$
$$= p_1(x)\varphi_1(x, \widehat{y}_1(x), \ldots, \widehat{y}_n(x), \widehat{u}_1(x), \ldots \widehat{u}_r(x)) + \ldots$$
$$\ldots + p_n(x)\varphi_n(x, \widehat{y}_1(x), \ldots, \widehat{y}_n(x), \widehat{u}_1(x), \ldots, \widehat{u}_r(x)).$$

Вот, наконец, и пришло время решить задачу Ньютона.

1°. Формализуем задачу Ньютона как задачу оптимального управления:

$$\int_0^a \frac{x\,dx}{1+u^2} \to \min, \quad y' = u, \ u \geqslant 0$$

с граничными условиями $y(0) = 0$, $y(a) = b$.

2°. Применяем принцип Лагранжа. Функция Лагранжа:

$$\mathscr{L} = \int_0^a \left(\frac{\lambda_0 x}{1+u^2} + p(y' - u) \right) dx.$$

Необходимое условие по y — уравнение Эйлера:

$$p' = 0 \implies p = \text{const} = p_0.$$

Необходимое условие по u — условие минимальности:

$$\frac{\lambda_0 x}{1+u^2} - p_0 u \geqslant \frac{\lambda_0 x}{1+\widehat{u}^2(x)} - p_0 \widehat{u}(x). \qquad (**)$$

3°. Если допустить, что $\lambda_0 = 0$, то необходимо, чтобы $p_0 \neq 0$ (иначе все множители Лагранжа оказались бы нулями). Если же $\lambda_0 = 0$ и $p_0 \neq 0$, то из (**) следует, что $\widehat{u} \equiv 0$, а значит, $y(x) = \int_0^x \widehat{u}(\alpha)d\alpha \equiv 0$. Тогда искомое тело «не имеет длины» — оно является плоской пластиной. Если же $b > 0$, то необходимо считать, что $\lambda_0 \neq 0$, и тогда можно положить $\lambda_0 = 1$. Отметим еще, что случай $p_0 \geqslant 0$ также невозможен, ибо при этом функция $\dfrac{x}{1+u^2} - p_0 u$ монотонно убывает и (**) невозможно.

Исследуя поведение функции $\dfrac{x}{1+u^2} - p_0 u = \varphi(u, x)$ по u, нетрудно убедиться в том, что при $p_0 < 0$ и малых x функция $\varphi(u, x)$ достигает минимума при $u = 0$. Затем оптимальное управление должно быть найдено из уравнения $-p_0 = 2ux/(1+u^2)^2$, получающегося из дифференцирования $\varphi(u, x)$ по u. Момент излома ξ определяется тем, что функция $\varphi(u, \xi)$ имеет два минимума.

Иначе говоря, в момент излома должны удовлетворяться соотношения (далее через $\widehat{u}(\xi)$ обозначается $\widehat{u}(\xi+0)\neq 0$):

$$-p_0=\frac{2\widehat{u}(\xi)\xi}{(1+\widehat{u}^2(\xi))^2},\quad \frac{\xi}{1+\widehat{u}^2(\xi)}-p_0\widehat{u}(\xi)=\xi.$$

Из второго уравнения получаем $-\xi\widehat{u}^2(\xi)/(1+\widehat{u}^2(\xi))=p_0\widehat{u}(\xi)$, откуда $p_0=-\xi\widehat{u}(\xi)/(1+\widehat{u}^2(\xi))$. Подставив это соотношение в первое из только что написанных уравнений, находим, что $\widehat{u}^2(\xi)=1\Longrightarrow$ $\Longrightarrow\widehat{u}(\xi)=1$, ибо $\widehat{u}\geqslant 0$. И тогда снова из того же уравнения находим, что $\xi=-2p_0$.

После излома оптимальное управление удовлетворяет соотношению

$$x=-\frac{p_0(1+u^2)^2}{2u}=-\frac{p_0}{2}\left(\frac{1}{u}+2u+u^3\right).$$

Но

$$\frac{dy}{dx}=u\Longrightarrow\frac{dy}{du}=\frac{dy}{dx}\frac{dx}{du}=u\frac{dx}{du}=-\frac{p_0}{2}\left(-\frac{1}{u}+2u+3u^3\right).$$

Интегрируя это соотношение, с учетом равенства $\widehat{y}(\xi)=0$ при $\widehat{u}(\xi)=1$, получаем параметрические уравнения кривой Ньютона

$$\widehat{y}(x,p)=\frac{p}{2}\left(\ln\frac{1}{u}+u^2+\frac{3}{4}u^4\right)-\frac{7p}{8},$$
$$x=\frac{p}{2}\left(\frac{1}{u}+2u+u^3\right),\quad p=-p_0.$$

Мы могли бы исследовать и простейшую задачу о быстродействии, но не будем делать этого. Отошлем читателя к книге [1], где эта задача решена с помощью принципа Лагранжа.

На этом поставим точку. Я выполнил свои обещания и все задачи первой части решил дважды. Давайте отдохнем от формул и поговорим просто так.

Итак, что же сказал бы я «первому встречному школьнику» о теории экстремальных задач (вспомним один из эпиграфов к этому рассказу)? Наверное, что-то в таком роде: в школе тебя учили функциям одной переменной. Там рассказывалось о приеме Ферма решения задач на экстремум таких функций. Но на самом деле имеется множество задач, сводящихся к минимизации функций многих переменных и даже функций от функции (например, от линии), так было, скажем, в задаче о брахистохроне. Их стали исследовать в специальном разделе — вариационном исчислении. В начале XX века была разработана научная дисциплина — функциональный (бесконечномерный) анализ, в котором обобщается основное понятие «школьного» анализа — понятие производной. Бесконечномерный анализ позволяет с единой точки зрения посмотреть на задачи минимизации функций одной и нескольких переменных и задачи вариационного исчисления.

При этом, если рассматривается задача без ограничений, то теорема Ферма полностью сохраняется: в точке экстремума эта производная должна равняться нулю. Для задач вариационного исчисления теорема Ферма расшифровывается в виде дифференциального уравнения, названного уравнением Эйлера.

Но задач без ограничений сравнительно мало. И среди задач с ограничениями большую долю составляют такие, что их можно формализовать в виде равенств.

Для конечномерных задач с равенствами Лагранж выдвинул принцип их решения. Суть его состоит в том, что надо составить функцию Лагранжа (т. е. просуммировать минимизируемую функцию и функции, задающие равенства с неопределенными множителями) и затем поступать с ней так, как будто ограничений нет. (Здесь я, может быть, просто процитировал бы слова Лагранжа, поставленные в качестве эпиграфа к двенадцатому рассказу.) Общий замысел Лагранжа сохраняется и для задач вариационного исчисления, и для задач оптимального управления — новой главы теории экстремальных задач.

А если бы мой новый знакомый проявил бы интерес к дальнейшему, я бы рассказал ему то, что написано во второй части этой книги. А следующий рассказ — о некоторых общих вопросах.

Рассказ пятнадцатый и последний, а на самом деле не рассказ, а беседа

Все жанры хороши, кроме скучного.

Французская пословица

Трудно не признать, что наши элементарные методы являются более простыми и более прямыми, чем методы анализа. Вообще, занимаясь той или иной научной проблемой, лучше исходить из ее индивидуальных особенностей, чем полагаться на общие методы.

Р. Курант и *Г. Роббинс*

В конце столетия существовала удручающая тенденция отворачиваться от основных проблем, как в механике, так и в чистом анализе. Вопреки великой традиции Якоба Бернулли и Эйлера, этот формализм быстро укрепился во французской школе и нашел отражение в «Аналитической механике».

Т. Трусделл

«Все жанры хороши, кроме скучного». В этой пословице отражен веселый и жизнерадостный дух французского народа. Как не хочется быть скучным!

В первой части этой книги я очень старался развлекать своего читателя. Я рассказывал легенды, притчи, истории и анекдоты, стремился быть разнообразным, наполнял содержание романтикой, выдумкой и поэзией, следовал за мыслями великих людей.

А во второй части все совершенно изменилось; стало приземленным, рутинным и прозаичным. Ни историй, ни поэзии, ни красот. Функции, производные, принцип Лагранжа... Сплошное однообразие! Формализация, необходимые условия, решение уравнений — и опять формализация, необходимые условия и т. п. Какая скука!

Вспомним еще раз один из наших эпиграфов. «Следует поставить перед собой цель изыскать способ решения всех задач одним и притом простым способом». (Даламбер имел здесь в виду задачи динамики.) Но возможно ли достигнуть такой цели? Разве неправда, что истины сокрыты в неисчислимом числе тайников? Можно ли ожидать, что они открываются одним ключом или хотя бы небольшой связкой ключей?

С Даламбером спорят Курант и Роббинс:

«Занимаясь той или иной научной проблемой, лучше исходить из ее индивидуальных особенностей, чем полагаться на общие методы». Известный современный механик Трусделл противопоставляет формализм Лагранжа великой традиции Эйлера, явно становясь на сторону Эйлера (см. эпиграфы к этой главе).

Кто же прав — Эйлер или Лагранж, Даламбер или Трусделл?

Не побоимся поставить и следующий вопрос, несмотря на кажущуюся однозначность ответа: что лучше — поэтическая романтика или скучное однообразие?

От этих вопросов не так уж легко отмахнуться, особенно тем, кто или учится, или учит. Чему учить и как учиться? Конкретному или абстрактному? Задачам или общим принципам?

Ограничим пока этим начальный круг проблем, которые предполагаем обсуждать.

Некоторые из поставленных вопросов могут показаться тривиальными. Начнем с самого простого — о романтике и рутине.

Романтика пленяет нас. Нас влекут горы, ледяные пустыни, бурные волны, опасность, риск. Мы чтим великих подвижников — путешественников, открывающих новые земли, альпинистов, взбирающихся на недоступные вершины, отважных и дерзновенных мореплавателей.

Когда-то достижение полюса было уделом героев. Но однажды мне довелось разговаривать с человеком, который побывал на обоих полюсах. В юности он любил путешествия, но потом у него случилось заболевание глаз, и он уже не мог носить большие рюкзаки и совершать дальние пешие переходы. И вот именно тогда-то он и побывал на полюсах. Для этого ему не пришлось ни мерзнуть, ни преодолевать торосы, ни проваливаться в полыньи. Он летал туда на самолетах.

И до сих пор можно пытаться дойти до полюса в одиночку или с группой друзей — на собаках, на лыжах, на дирижабле, с преодолением препятствий, с романтикой и риском.

Но есть и другой путь к цели — самолетом. Когда он начался? Не тогда ли, когда было изобретено колесо? Потом телега, паровой двигатель, железная дорога, автомобиль и, наконец, самолет. Самолет может доставить на полюс любого из нас — без поэзии и романтики.

Человечество не может двигаться вперед без подвижничества, без одержимости, без риска, без романтики и поэзии. Но неизбежно наступает время, когда манившая и далекая цель становится общедоступной, освоенной. Тогда для достижения ее никто не делает ничего героического. Кто-то готовит самолет к вылету, совершая обычную, ежедневную работу. Потом приходит команда. На аэродроме нет провожающих, нет оркестров. Люди просто выполняют свою работу. Самолет вылетает и потом приземляется на полюсе. Но лишь тогда, когда такой полет становится рядовым, рутинным, обычным, заурядным, — лишь тогда можно сказать, что полюс оказался освоенным.

Разве кто-нибудь может усомниться в том, что в будущем, когда человечество справится с бременем отягчающих его проблем, оно не позаботится о том, чтобы любой человек мог постоять на Эвересте, не карабкаясь и не задыхаясь?

Прогресс в жизни и в науке осуществляется соединением усилий первопроходцев и размеренным движением вперед всей нашей цивилизации. Это движение со временем делает общедоступными те цели, к которым ценою жертв и страданий раньше пробирались лишь герои. Как и в жизни, в науке каждый человек может ставить перед собой двоякие цели. Он должен тренировать себя для того, чтобы, даже не располагая особыми средствами, сделать попытку в одиночку взобраться на неприступную вершину. Но нужно также участвовать в коллективном труде, обеспечивающем размеренное движение цивилизации по прокладыванию дорог и коммуникаций к этим вершинам. И следовательно, учить тоже и тому и другому.

Вот мы подошли к теме — чему учить и как учить. Учатся для развития (в основном — до школы), для образования (в школе), для профессии (в институте и аспирантуре). О каждом этапе нужно думать отдельно. Могут спросить: а разве это уже не продумано? Ведь это — один из самых кардинальных вопросов, касающийся и каждого отдельного человека, и всего человечества.

Когда-то мне думалось, что все вопросы уже решены и все давно известно — на земле, в небе, в науке и в жизни. Ведь жизнь на Земле длится так долго, столько было мудрецов!

В этой книге я старался представить один из разделов математики — от истоков до наших дней. Окинем взором пройденное еще раз. Давно ли все начиналось? И да, и нет. Двадцать пять веков... Это, конечно, очень много. Но выстроим мысленный ряд, по одному человеку от каждого поколения, пусть пять человек в век. Окажется сто двадцать пять человек — от нас до Аристотеля. Как мало!

Мы еще очень молоды. Человечество только еще начинает осваивать мир. Собственно, наука о природе в нынешнем понимании этого слова зародилась лишь триста-четыреста лет назад. И уже так много известно. И так мало! Ваши дедушки и бабушки могли знать людей, родившихся в век телеги. Буквально на глазах, в течение немногим более ста лет человечество освоило практически все, что наполняет сейчас нашу жизнь — пароходы, железные дороги, телеграф, телефон, автомобиль, самолет, телевизор, спутник. У человечества еще не было времени, чтобы в деталях продумать многие наиважнейшие жизненные проблемы. В частности, не пришли к полной ясности по вопросу о том, чему учить... Но вне зависимости от того, как в будущем (и в наши дни) решится вопрос о содержании образования (а здесь можно дискутировать о многом — не следует ли с самых ранних пор обучать наряду с обычными предметами общению с ЭВМ, вождению автомобиля, стенографии и машинописи, искусству редактирования и т. п.), для меня нет сомнения в том, что математике учить нужно по меньшей мере, по двум причинам — для тренировки ума и для того, чтобы можно было понимать, как устроен мир.

Дискуссия о том, как учить математике, не утихает на протяжении последних ста лет. В начале в ней принимали участие все крупнейшие ученые — Клейн, Борель, Адамар и др. Для того чтобы обозначить один из предметов этого неутихающего диспута, приведу большую цитату из Дьедонне — одного из самых видных французских математиков нашего времени:

«Я прошу всех беспристрастно посмотреть на следующие темы, занимающие большое место в школьной математике:

I. Задачи на построение «циркулем и линейкой».

II. Свойства «традиционных» фигур, таких, как треугольники, четырехугольники, окружности и системы окружностей... — все это со всеми изощрениями, накопленными поколениями «геометров» и преподавателей в поисках подходящих экзаменационных задач.

III. Весь псалтырь «тригонометрических формул» и их калейдоскопических преобразований, позволяющих находить великолепные «решения задач» на треугольники и — пожалуйста, имейте это в виду — «в форме, пригодной для логарифмирования...»

Далее Дьедонне пишет, что ни с чем подобным человек никогда в жизни не столкнется. С тем же темпераментом и напором он низвергает «старую школьную математику»: «Но спрашивается, важнее ли при этом для конструктора знать, что высоты пересекаются в одной точке, или же овладеть принципами теории сопротивления материалов?»

Дьедонне считает, что надо учить принципам и только им! (Прочтите введение к книге Ж. Дьедонне «Линейная алгебра и элементарная геометрия» (М.: Наука, 1972) там вы найдете еще много интересного.)

Я обучался «по-старому» и к перечню Дьедонне могу добавить еще многое, например «арифметические задачи» типа задач на бассейны, бесконечные арифметические примеры, где обычные дроби складываются с десятичными, и т. п.

И снова может показаться, что нет предмета для дискуссий, что Дьедонне прав. Ведь это действительно так, и «тригонометрические формулы совершенно необходимы для представителей трех очень почтенных профессий: для астрономов, для геодезистов, для составителей учебников тригонометрии, а больше никому. Так зачем же ими морочить голову бедному школьнику?

И все-таки что-то мешает мне признать правоту этих слов. Само понятие «образование» более сложно. Оно состоит не только в приобретении знаний и навыков, но и в тренировке мышления. На протяжении, по-видимому, двух столетий (а может быть, и больше) задачи на бассейны, задачи на построение, задачи на треугольники и преобразования тригонометрических формул выполняли великую роль — они давали пищу для ума, приучали к точности и аккуратности, учили рассуждать, искать истину, преодолевать трудности, испытывать разные пути к цели, учили достигать ее. Они одаривали радостью успеха и ощущением красоты. В конечном счете, они *моделировали творчество*. Чем заменить все это? И стоит ли?

Все эти элементы творчества, необходимо, конечно, сохранить. Быть может, на другом материале, но сохранить. *Тренировать мышление можно лишь на конкретных, «частных» задачах*, а не на «общих принципах». Мне кажется, что надо дать возможность решить задачу Герона по Герону, на задачу Евклида посмотреть глазами Евклида, дать почувствовать сопротивление задачи Архимеда в тот момент, когда вы достигли уровня его технических средств, попробовать самостоятельно решить задачу Штейнера.

Мне представляется, словом, что среди тех тем, на которых можно тренировать мысль, приучать к изобретательности, научной изворотливости, преодолению интеллектуальных трудностей, задачи на экстремум представляют благодарный материал. Поэтому я и писал первую часть этой книги.

Но несомненно и то, что надо обучать пониманию сущности вещей, общим принципам и законам — и в естествознании, и в жизни. Для этого необходимо, как мне кажется, хотя бы в самых общих чертах познакомить любого человека с основами математического анализа. Ибо математический анализ — неотделимая часть естествознания. Но важно еще и другое. Важно, чтобы были как-то осознаны и поняты и единство мира, и его разнообразие. Электричество, свет, тепло, движение жидкости, движение планет — это все разное, но в этом разном имеются общие черты. Природа «управляется» общими законами, все связано и все тяготеет к единству. Мне кажется, это тоже важно — знать, что существуют «общие принципы». Такие принципы бывают и в математике. И Лагранж, и Даламбер, и многие другие великие ученые старались возвыситься до понимания той «самой сути», которая соединяет разрозненные явления в мире. Об одном таком общем принципе — принципе Лагранжа — было рассказано во второй части этой книги.

Вспомним: каждую задачу мы старались решить дважды. Один раз — в первой части, «исходя из ее индивидуальных особенностей», другой раз, — во второй, «полагаясь на общие методы», я напоминаю еще раз слова Куранта и Роббинса. У них получалось, что «элементарные методы являются более простыми и прямыми, чем методы анализа». А у нас?

Мне кажется, что общий метод Ферма—Лагранжа, по меньшей мере, не уступает первенства в этом споре. Конечно, он иногда терпит и поражения, скажем, в задаче о наименьшем периметре треугольника (возможно, впрочем, мое решение не оптимально). Не исключено, что кто-то из моих читателей присудит победу другим изящным элементарным решениям. Но разве можно отрицать, что этот метод сражался достойно, ни разу не отказавшись от единоборства, и всегда приводил нас к цели. А в скольких случаях он одерживал уверенную победу! Это может показаться удивительным, ведь один в поле не воин! Но можно усмотреть в этом определенную закономерность, сопутствующую многим человеческим исканиям. Проблеск истины рождается во мраке, человек долго блуждает, пробираясь сквозь дебри, а потом вдруг обнаруживается, что все эти долгие мучительные поиски были как бы напрасны и к цели ведет короткая дорога. Но напрасны ли? Подумайте над этим сами. Мне хотелось своей книгой дать вам возможность проследить за тернистой дорогой поисков истины.

И еще один аспект высказывания Куранта и Роббинса хотелось бы обсудить. В те годы, когда писалась книга Куранта и Роббинса, слишком много значения придавалось противопоставлению элементарной и высшей математики. Самому этому противопоставлению триста лет. Границы элементарной математики определяются соседством с математическим анализом. Все, что было создано до рождения математического анализа, да и многое из того, что появилось позже, но не использует его методов и конструкций, принято относить к элементарной математике. Математический же анализ иногда называют «высшей математикой». Некогда считалось, что в высшей математике содержится нечто «сверхъестественное», недоступное пониманию простого человека, нечто воистину «высшее», о чем в школе и говорить невозможно. Но на самом деле это не так.

Математический анализ — это совершенно естественная, простая и элементарная наука, ничуть не более заумная, сложная или «высшая», чем, скажем, «элементарная» геометрия. Многие теоремы, традиционно входившие в курс геометрии, куда сложнее, чем основополагающие теоремы классического анализа. Ныне противопоставление элементарной математики и анализа непродуктивно, и вовсе необязательно проявлять бездну остроумия только лишь из боязни использовать свойства производной.

Привнесение элементов математического анализа в школьные программы неизбежно приведет к перестройке и других областей математического образования — изменится содержание конкурсных задач, кружковой работы, математических олимпиад и многого другого. Теперь уже невозможно не учитывать того, что школьник должен знать нечто из ранее недоступной ему высшей математики. Возможности, которые скрыты в самых простых средствах математического анализа, я старался продемонстрировать в тринадцатом рассказе, когда мы заново решали задачи из первой части и обсуждали некоторые вопросы, которым там (на элементарном уровне) не нашлось места.

При этом следует иметь в виду, что если освоены лишь самые основы математического анализа, можно уже делать попытки подобраться ко многим современным проблемам. В четырнадцатом рассказе я постарался осветить тот путь, который был пройден в теории экстремальных задач от Ньютона, Лейбница, Эйлера и Лагранжа до наших дней. Мне хотелось показать, как, в сущности, недалёко от Ньютона до нас.

Эпиграфом к первой части этой книги были избраны слова Бертрана Рассела, где математика сопоставляется с искусством. Перед началом четвертого рассказа были приведены слова Г. Харди, где он Ученого ставит выше Поэта. Не будем вступать в спор. И наука, и искусство соединяются в общем понятии — достояние человеческого разума. И мне хотелось, чтобы мой читатель воспринял историю небольшого фрагмента математической науки как часть этого нашего общего культурного достояния.

Список литературы

[1] Алексеев В. М., Тихомиров В. М., Фомин С. В. Оптимальное управление. — М.: Физматлит, 2005.

[2] Алексеев В. М., Галеев Э. М., Тихомиров В. М. Сборник задач по оптимизации. — М.: Физматлит, 2005.

[3] Белый Ю. А. Иоганн Кеплер. — М.: Наука, 1971.

[4] Бляшке В. Круг и шар. — М.: Наука, 1967.

[5] Бляшке В. Греческая и наглядная геометрия // Математическое просвещение. — М.: Физматгиз, 1958.

[6] Болтянский В. Г., Яглом И. М. Геометрические задачи на максимум и минимум // Энциклопедия элементарной математики, кн. V. — М.: Наука, 1966. С. 270—348.

[7] Гиндикин С. Г. Рассказы о физиках и математиках. — М.: МЦНМО, 2006.

[8] Декарт Р. Геометрия. — М.—Л.: ГИТТЛ, 1938.

[9] Зетель С. И. Задачи на максимум и минимум. — М.—Л.: Гостехиздат, 1948.

[10] Кокстер Г. Введение в геометрию. — М.: Наука, 1966.

[11] Кеплер И. Стереометрия винных бочек. — М.—Л.: Гостехиздат, 1935.

[12] Крыжановский Д. А. Изопериметры. — М.—Л.: ОНТИ, 1938.

[13] Курант Р., Роббинс Г. Что такое математика? — М.: МЦНМО, 2004.

[14] Люстерник Л. А. Выпуклые фигуры и многогранники. — М.: Гостехиздат, 1956.

[15] Ньютон И. Математические работы. — М.—Л.: ГИТТЛ, 1937.

[16] Предтеченский Е. А. Кеплер. Его научная жизнь и деятельность. — Петроград: Изд-во Гржебина, 1921.

[17] Радемахер Г., Теплиц О. Числа и фигуры. — М.: Физматгиз, 1962.

[18] Тарасов Л. В., Тарасова А. Н. Беседы о преломлении света. — М.: Наука, 1982. — Библиотечка «Квант», вып. 18.

[19] Харди Г., Литлвуд Дж., Полиа Г. Неравенства. — М.: УРСС, 2006.

[20] Цейтен Г. Г. История математики в древности и в средние века. М.—Л.: ГТТИ, 1932.

[21] Шарыгин И. Ф. Задачи по геометрии. Планиметрия, изд. 2-е. М.: Наука, 1986. — (Библиотечка «Квант», вып. 17.)

[22] Шарыгин И. Ф. Задачи по геометрии. Стереометрия. — М.: Наука, 1984. — (Библиотечка «Квант», вып. 31.)

[23] Шклярский Д. О., Ченцов Н. Н., Яглом И. М. Геометрические неравенства и задачи на максимум и минимум. — М.: Наука, 1970.

[24] Яглом И. М., Болтянский В. Г. Выпуклые фигуры. М. — Л.: Гостехиздат, 1951.