Д. В. АНОСОВ

Отображения окружности, векторные поля и их применения

МЦНМО
Москва 2003

УДК 515.12
ББК 22.152
А69

Аносов Д. В.

А69 Отображения окружности, векторные поля и их применения. — М.: МЦНМО, 2003. — 120 с.

ISBN 5-94057-101-8

В книге доказывается теорема Жордана. С этой целью вводятся основные топологические понятия (степень отображения, векторные поля) и доказываются многие топологические утверждения (основная теорема алгебры, теорема о числе вещественных корней многочлена, критерий Эйленберга и др.).

Для школьников и студентов физико-математических специальностей.

ББК 22.152

Дмитрий Викторович Аносов

Отображения окружности, векторные поля и их применения

Издательство Московского центра
непрерывного математического образования

Лицензия ИД № 01335 от 24.03.2000 г.

Подписано в печать 03.09.2003 г. Формат 60×90 $^1/_{16}$. Печ. л. 7.5.
Тираж 1000 экз. Заказ № .

МЦНМО
119002, Москва, Большой Власьевский пер., 11

Отпечатано во ФГУП «Полиграфические ресурсы».

ISBN 5-94057-101-8

Предисловие

Эта книжка призвана дать некоторое понятие о топологии — относительно новом разделе математики, оформившимся в самостоятельную дисциплину только в конце XIX в. Некоторый намёк на её содержание даётся во введении.

В книжке подразумеваются известными некоторые элементы математического анализа, в основном те, которые связаны с непрерывностью и пределами; в § 2 небольшую роль играют также и производные. Студенты-математики или студенты, будущая профессия которых требует более или менее солидного изучения математики, обычно проходят всё это на первом семестре первого курса. С данным материалом бывают знакомы и хорошо подготовленные школьники, особенно учащиеся в физико-математических школах или классах.

Прежде чем рассказывать собственно о топологии, а точнее, о некоторой её части, пришлось изложить различный вспомогательный материал, который отчасти очень важен сам по себе, но не относится к топологии (§ 2), а отчасти относится, но не к той её части, ознакомление с которой является нашей целью (§ 3). Вспомогательные §§ 2, 3 получились довольно длинными. Я исходил из того, что читателю, скорее всего, многое здесь уже известно (если он — студент, то это заведомо так), но известно всё же не всё и, возможно, материал у него в голове ещё не «отложился» в достаточно чётком виде, причём здесь я не мог считать, что такие-то вещи ему хорошо известны, а такие-то — нет, потому что у различных читателей дела тут наверняка обстоят по-разному. В таком случае какую-то часть §§ 2, 3 ему придётся прочитать внимательно, а значение другой части для него, самое бóльшее, сведётся к тому, чтобы что-то уточнить, восполнить небольшие пробелы в его знаниях и помочь «разложить по полочкам» этот материал.

§ 1. Введение

Топология имеет дело с такими свойствами геометрических фигур, которые не меняются, если фигуры деформировать (считая их, так сказать, резиновыми), но не разрывать и не надрывать. То, что такие свойства существуют, понятно. Например, ясно, что кривая может быть замкнутой или нет, и это ни при какой её непрерывной деформации не изменится; наглядно не вызывает сомнений, что сферу невозможно продеформировать в тор (поверхность баранки, см. рис. 1), хотя доказать последнее непросто.

Однако далеко не столь ясно, что возможно более или менее систематически исследовать такие свойства, и что это может найти приложения

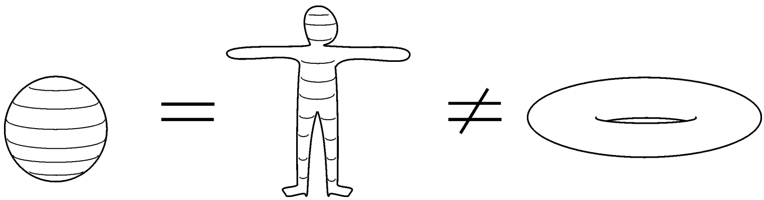


Рис. 1

в других разделах математики (а теперь также и в физике). Не случайно топология как самостоятельная дисциплина сложилась довольно поздно.

Топология — это, так сказать, геометрия непрерывности, одной только непрерывности, когда игнорируется всё остальное — расстояния, углы, прямолинейность и т. д. Уже на доматематическом интуитивном уровне представляется наглядно ясным различие между непрерывностью и разрывностью. Одно из первых, что делают в курсе математического анализа — это уточняют идею непрерывности (хотя поначалу и не в полной общности), давая чёткое формальное определение непрерывности функции (как и родственного понятия предела) и доказывая несколько примыкающих сюда утверждений. Однако основные объекты математического анализа — производные, ряды и интегралы — связаны преимущественно с другими идеями, а непрерывность (да в какой-то степени и предел) — не более чем необходимый пререквизит. Насколько он необходим для курса анализа, настолько он там и развивается, но не более того. В топологии же идея непрерывности выступает на первый план.

Сейчас я уточню свойственную топологии точку зрения. Фигуры A и B (для начала они могут быть подмножествами плоскости или обычного пространства) называются гомеоморфными, если существует непрерывное взаимно однозначное отображение[1)] одной фигуры на другую, обратное к которому тоже непрерывно. Такое отображение называется гомеоморфизмом. Можно в первом приближении считать, что топология занимается свойствами фигур, не меняющимися при гомеоморфизмах, так что для неё гомеоморфные фигуры одинаковы. Такие свойства называют топологическими свойствами. Ясно, что в предыдущих двух примерах (замкнутость кривой; тор $\neq$ сфера) мы имеем дело именно с топологическими свойствами. Понятно также, почему замкнутую кривую без самопересечений называют топологической окружностью, а поверхность, гомеоморфную сфере, — топологической сферой.

Сказанное даёт не более чем намёк на содержание топологии, потому что хотя и сформулирована свойственная ей точка зрения, но, повторяю,

---

[1)]Напомню, что отображение $f\colon A \to B$ (пишут также $A \xrightarrow{f} B$) множества A («области определения f»; сейчас это не обязательно геометрическая фигура) в множество B, или функция, определённая на A и принимающая значения в B, — это соответствие, при котором каждому элементу $x \in A$ сопоставляется некоторый элемент из B. Последний элемент обозначают через $f(x)$ и называют образом элемента x при отображении f, или значением функции f в элементе x; пишут $x \mapsto f(x)$. Образом $f(C)$ подмножества $C \subset A$ называется совокупность элементов $f(x)$ со всевозможными $x \in C$ (так сказать, то, что получается при данном отображении из C).

Когда $f(A) = B$ (т. е. каждый элемент из B является образом некоторого элемента из A; «под действием» f множество A как бы покрывает всё B), то говорят, что f отображает A на B. Если же не утверждается, что $f(A) = B$, то говорят об отображении A в B. Отображение $f\colon A \to B$ называется взаимно однозначным отображением A на B, если образы различных элементов из A всегда различны (и, коль скоро сказано «на B» — каждый элемент из B является образом некоторого элемента из A). Для такого f определено обратное отображение $f^{-1}\colon B \to A$, сопоставляющее каждому элементу $y \in B$ тот элемент из A, образом которого при исходном отображении f является y. (В тех случаях, когда значения f суть ненулевые числа, имеется некоторая опасность путаницы, потому что под $f^{-1}(x)$ можно понимать также и число $\dfrac{1}{f(x)}$. В связи с этим было предложено обозначать обратное отображение через $\overset{-1}{f}$. Однако опыт показывает, что при достаточном внимании к контексту путаницы не возникает.)

Ещё два замечания в связи с обозначениями и терминологией. Выше использованы две стрелки: $\to$ и $\mapsto$. Их надо различать. Первая (оставляя в стороне её использование для обозначения сходимости) употребляется, когда надо указать, что одно множество является областью определения данного отображения и что это множество отображается в другое. Вторая стрелка применяется в записи типа $x \mapsto f(x)$, указывая, что элементу x (непосредственно не сказано, какому множеству он принадлежит, — это подразумевается известным) сопоставляется элемент $f(x)$ (тоже не сказано, какого множества).

Отображения и функции — это, собственно, синонимы, но первый термин возник в геометрии, а второй — в математическом анализе. Это до некоторой степени отражается в употреблении данных терминов. Функции чаще всего принимают числовые значения или, во всяком случае, такие значения (скажем, векторные), над которыми можно производить какие-то алгебраические операции. Значения, принимаемые отображениями, чаще, чем значения функций, бывают элементами каких-то множеств, где ни о каких алгебраических операциях говорить не приходится.

едва ли заранее ясно, что можно более или менее систематически исследовать свойства фигур, сохраняющиеся при гомеоморфизмах (да и так ли уж много подобного рода свойств и окажутся ли они интересными?). Конечно, можно привести ещё несколько наглядных примеров. Скажем, проделав в круге[2] n не соприкасающихся друг с другом дыр, получаем при одном и том же n топологически одинаковые фигуры, независимо от расположения и размеров этих дыр, а при различных n — не гомеоморфные друг другу фигуры. Но это не очень помогает. К сожалению, дать более чем намёк, ограничиваясь несколькими фразами или даже несколькими абзацами, едва ли возможно. В популярной книжке [4] правильно говорится:

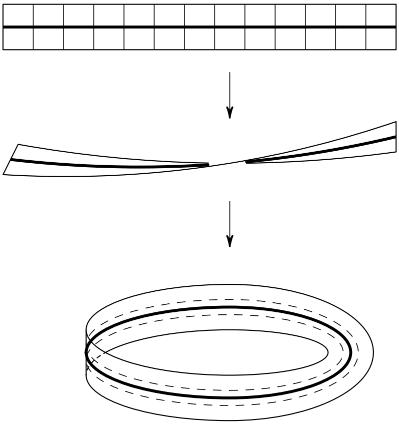


Рис. 2. Лист Мёбиуса

«Когда нематематик спрашивает тополога: „Что такое топология?“, „Для чего она нужна?“, последний оказывается в невыгодном положении. Спрашивающий ждёт ответа, который можно дать на аналогичные вопросы относительно тригонометрии, вроде того, что тригонометрия имеет дело с определением углов и используется для решения задач в геодезии, навигации и астрономии. Но такого прямого ответа тополог дать не может. Он может с полным основанием сказать, что топология — это род геометрических соображений, полезных во многих областях современной математики. Однако это не удовлетворит того, кто хотел бы уловить суть предмета. Тогда тополог может принести сделанный из бумаги с помощью ножниц и клея лист Мёбиуса и разрезать его вдоль центральной линии[3] или же он может взять верёвку и показать, как мож-

[2] При нашей уточнённой терминологии круг отличается от окружности так же, как шар от сферы (такое различие не всегда соблюдают). Круг состоит из всех тех точек плоскости, расстояние которых до одной и той же её точки (центра) не превосходит r, а окружность — из тех, расстояние которых до центра равно r.

[3] Лист Мёбиуса (рис. 2) — популярная тема популярной математической литературы. Если читатель ещё не имел с ним дела, советую срочно склеить несколько листов Мёбиуса и проделать следующие опыты. 1) Разрежьте его по его средней линии (жирная линия на рисунке). 2) Отрежьте от него кусок, отстоящий от края листа Мёбиуса на треть его ширины (разрез изображён пунктиром). 3) Попробуйте закрасить его с одной стороны. 4) Выясните, сколько у него краёв (т. е. ограничивающих его линий). Хотя авторы [4] относятся к подобной деятельности с резонным скептицизмом, эти опыты всё же полезны в одном отношении:

но, не связывая её, образовать три отдельные петли [4]. Если он почувствует прилив энергии, он может продемонстрировать, как снять пиджак, не снимая пальто [5]. Каждый из этих домашних фокусов основан на серьёзной математической идее, требующей для своего объяснения по крайней мере нескольких часов. Но показывать эти фокусы без соответствующего объяснения значило бы изображать карикатуру на топологию.»

Резонно отказавшись от подобных фокусов, авторы [4] посвятили свою книжку аккуратному изложению некоторого «настоящего», и в то же время элементарного, «кусочка топологии». Этот «кусочек» связан с тем же «техническим аппаратом», что и в настоящей книжке, но в [4] этот аппарат строится несколько иначе (и служит отчасти для других целей). В предварительном порядке можно уже сейчас сделать некоторые пояснения об этом аппарате.

Напомню сперва элементарный (но полезный) факт из самого начального курса анализа. Пусть на отрезке $[a, b]$ задана непрерывная функция $f\colon [a, b] \to \mathbb{R}$ (принимающая, как видно, обычные числовые значения [6]). Что можно сказать о её «нулях» (т. е. о решениях уравнения $f(x) = 0$) на основании информации о значениях f на концах отрезка $[a, b]$? Мы считаем, конечно, что эти значения отличны от нуля — в противном случае у нас уже есть нуль. Сказать можно не так уж много, но кое-что всё-таки можно: если $f(a)$ и $f(b)$ имеют противоположные знаки, то f обязательно обращается в нуль где-то внутри данного отрезка. (Точно знать $f(a)$ и $f(b)$ при этом даже не требуется.) Если же знаки $f(a)$ и $f(b)$ одинаковы, то ничего сказать нельзя — в этом случае f не обязана обращаться в нуль на отрезке $[a, b]$ (например, это так для линейной функции

они показывают, что даже совсем простые геометрические фигуры могут иметь непривычные свойства.

[4] По-видимому, имеется в виду, что после этого концы верёвки соединят и туго завяжут друг с другом, так что получится замкнутая кривая с узлами. Она гомеоморфна замкнутой кривой без узлов, но иначе расположена в пространстве. Это тоже относится к ведению топологии при разумном понимании «сохранения свойств при гомеоморфизме»: надо говорить не о гомеоморфизме одной кривой на другую, но о гомеоморфизме всего пространства, при котором одна кривая переходит в другую. Если одна замкнутая кривая заузлена, а другая — нет, то такого гомеоморфизма не существует. — *Д. А.*

[5] Легенда гласит, что нечто в этом роде публично проделал выдающийся английский тополог Дж. Г. К. Уайтхед, когда в связи со своим избранием президентом Королевского Математического Общества он читал лекцию о своей области науки. Только он извлёк свитер из-под пиджака (что свидетельствовало как о его ловкости, так и о качестве шерсти). Возможность извлечения пиджака из-под пальто — это не чисто топологический вопрос, ибо материал пиджака почти не растягивается (а только легко изгибается). Хотя строгое математическое исследования последней возможности едва ли проводилось, в ней позволительно усомниться. — *Д. А.*

[6] $\mathbb{R}$ — это более или менее стандартное обозначение для множества вещественных чисел (тогда как обычная буква R может, по мере надобности, обозначать что угодно).

со значениями одного знака на концах отрезка), но, конечно, может и иметь там нуль.

А теперь зададимся аналогичным вопросом о системе двух уравнений с двумя неизвестными. Пусть у нас имеются две функции от двух переменных $f(x, y)$, $g(x, y)$, заданные и непрерывные на круге $\{(x, y);\, x^2+y^2 = r^2\}$. Имеется ли какое-нибудь условие об их поведении на ограничивающей этот круг окружности[7] $\mathbb{S}^1_r = \{(x, y);\, x^2 + y^2 = r^2\}$, гарантирующее, что внутри круга обязательно найдётся точка (x, y), где и f, и g обращаются в нуль? Да. Его формулировка как раз и связана с теми топологическими понятиями, которым посвящена эта книжка. Наглядно его можно описать так. Рассмотрим вектор $\boldsymbol{v}(x, y)$ с координатами $f(x, y)$, $g(x, y)$. Аналогично предыдущему, мы считаем, что нигде на $\mathbb{S}^1_r$ он не равен нулю, — в противном случае мы уже имеем точку, где $f = g = 0$. Представим себе, что точка с координатами (x, y) делает один оборот по окружности $\mathbb{S}^1_r$ против часовой стрелки, и будем следить, как при этом меняется направление вектора $\boldsymbol{v}(x, y)$. Сейчас удобнее представлять себе, что этот вектор приложен к началу координат $O = (0, 0)$. Когда (x, y) делает оборот по $\mathbb{S}^1_r$, вектор $\boldsymbol{v}(x, y)$ в свою очередь делает несколько оборотов вокруг O. («Петля», когда $\boldsymbol{v}$ сперва идёт в одну сторону, а затем возвращается назад, за оборот не считается.) Оказывается, что если число оборотов отлично от 0, то внутри нашего круга обязательно имеется точка, где $\boldsymbol{v} = 0$. Если же оно равно нулю, то ничего сказать нельзя — такая точка может не существовать, а может и существовать.

Возможно, читатель слышал о такой теореме о неподвижной точке: при непрерывном отображении F замкнутого круга в себя обязательно имеется точка, которая неподвижна, т. е. переходит в себя. Эта теорема легко доказывается применением предыдущего утверждения к «вектору смещения», соединяющему каждую точку z круга с её образом $F(z)$. Дело в том, что при обходе z по $\mathbb{S}^1_r$ этот вектор делает один оборот вокруг O. А точка с нулевым вектором смещения — это и есть неподвижная точка.

В популярной литературе, как только речь заходит о топологии, обычно сразу же приводят теорему о неподвижной точке. Я, как видно, этого тоже не избежал. Но всё-таки я не буду её доказывать. Доказательства, основанные на той же идее, которая намечена выше, можно найти в книгах, указанных в списке литературы.

Я говорил вначале о гомеоморфных фигурах, а теперь перешёл к свойствам непрерывных отображений. Это, пожалуй, самое важное уточне-

---

[7] $\mathbb{S}^1$ — это более или менее стандартное обозначение для окружности единичного радиуса на плоскости с центром в начале координат. Обозначение $\mathbb{S}^1_r$ для окружности с тем же центром, но радиуса r, является, может быть, чуть менее стандартным, но в свете предыдущего обозначения оно представляется вполне естественным.

ние, которое надо внести в данную ранее приблизительную характеристику предмета топологии — топологические свойства непрерывных отображений[8)] интересуют топологию не меньше, чем топологические свойства фигур (на самом деле вторые нередко изучают с помощью первых).

Давая аккуратное определение понятий типа «числа оборотов $\boldsymbol{v}$ вокруг точки O» и устанавливая некоторые их свойства, мы фактически строим некоторый, хотя и небольшой, технический аппарат. В отличие от [4], при построении этого аппарата я использую комплексные числа, которые очень важны совершенно независимо от такого их использования, поскольку играют большую роль во многих вопросах совсем иного характера. Возможно, читатель в какой-то степени знаком с комплексными числами, но на всякий случай я напоминаю то, что нам нужно, в начале § 2. Основная же часть этого параграфа посвящена экспоненциальной функции в комплексной области. Для слабо подготовленных читателей имеется резюме, так сказать «сухой остаток», где объясняется, что на самом деле из всего § 2 формально нам достаточно очень немногого. Я мог бы с самого начала ограничиться этим немногим, но экспонента в комплексной области — это тема, которая сама по себе настолько интересна и важна, что жалко ограничиваться какими-то её обрывками.

Как и в настоящей книге, в [4] предварительно обсуждаются общие понятия и факты, связанные с непрерывностью. (В [4] это делается подробнее, но нет теоремы о продолжении функции, которая там не нужна.) В настоящей книжке этому посвящён § 3. На «высоком штиле», его содержание относится к «общей (или теоретико-множественной) топологии», но, разумеется, при изложении данного материала я имею в виду нужную нам ситуацию и соответственно стараюсь упростить это изложение. Как уже говорилось, вспомогательные §§ 2, 3 получились довольно длинными; можно ещё добавить, что при всём том в § 3 маловато примеров, хотя содержательное усвоение соответствующего материала требовало бы таковых (причём они должны быть достаточно разнообразными). Повторяю, что читателю, скорее всего, многое здесь уже известно, но различным читателям наверняка известны различные вещи и в различной степени; поэтому мне пришлось написать обо всём, что потребуется далее, а читатель сам решит, что из этого и насколько внимательно ему надо прочитать.

[8)] Что бы это ни значило — я ведь не говорил, какие свойства непрерывных отображений называются топологическими. Мы будем иметь дело только с теми свойствами непрерывных отображений, которые не меняются при непрерывных деформациях этих отображений. Точное определение непрерывной деформации отображения даётся в § 4, но уже сейчас можно представить себе, что если можно непрерывно деформировать резиновую фигуру, то можно непрерывно деформировать и отображение.

Вместо теоремы о неподвижной точке я предпочёл привести два приложения излагаемых здесь топологических понятий к алгебре (§ 6). Первое из них тоже является излюбленным сюжетом популярной литературы, зато второе не только в ней не встречается, но, возможно, вообще является новым. Я узнал о нём от его автора, А. Г. Хованского, которому я благодарен за это сообщение, сделанное в самое подходящее для меня время — когда я думал, какое бы новое применение элементарной топологии я бы мог привести.

Конец настоящей книжки посвящён доказательству теоремы Жордана, утверждающей, что замкнутая кривая L на плоскости Π, не имеющая самопересечений, разбивает плоскость на две части (см. рис. 3).

Это значит, что дополнение к L на плоскости, т. е. $\Pi \setminus L$, состоит ровно из двух связных частей. Данное утверждение на первый взгляд представляется совершенно очевидным, а намерение дать ему строгое доказательство кажется ненужным педантизмом. Поэтому уместно прежде всего разъяснить, почему интуитивно несомненная теорема Жордана далеко не очевидна.

Рис. 3

Интуитивная убеждённость в том, что любая непрерывная замкнутая кривая без самопересечений разбивает плоскость, основано в конечном счёте на опыте обращения с простыми линиями, вроде окружности, эллипса, периметра треугольника или параллелограмма, и т. п. Во всех этих случаях достаточно одного взгляда, чтобы установить, какие точки лежат внутри кривой и какие — снаружи. В подобных случаях не может составить труда и проведение строгого доказательства «по всей форме», если возникнет желание иметь таковое (причём здесь это действительно отдаёт чрезмерным педантизмом), потому что каждый раз легко указать условие, отличающее те точки, которые лежат внутри кривой, от тех, которые лежат снаружи (например, в случае окружности расстояние от «внутренних» точек до центра меньше радиуса окружности, а от «внешних» — больше).

Однако в случае какой-нибудь извилистой замкнутой кривой (как на рис. 4) нужно некоторое время всматриваться в чертёж, чтобы понять, лежит ли та или иная точка внутри или снаружи кривой, так что едва ли можно назвать утверждение, что эта кривая разбивает плоскость, таким уж очевидным. Таковым оно становится лишь после того, как удастся заштриховать на рисунке или наглядно выделить в уме ту область, которая ограничена нашей кривой. Штриховка представляет собой эмпирическую проверку теоремы Жордана для данного случая, но ведь нельзя же перебрать все

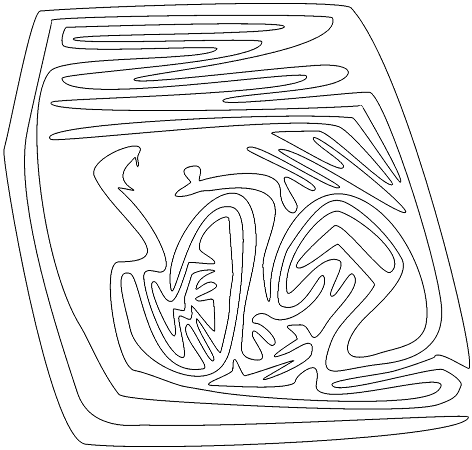


Рис. 4

замкнутые кривые, какие только могут быть, и для всех них тем же способом проверить теорему Жордана. Общего подхода подобная эмпирическая деятельность не даёт. И ведь известно, что усложнение формы кривой может быть связано с её локальным строением, которое вообще невозможно точно изобразить на рисунке. Наиболее известный пример — непрерывная кривая, ни в одной своей точке не имеющая касательной; менее известно, что существуют кривые, которые, не имея самопересечений, «ухитряются» тем не менее иметь положительную площадь (см., например, [6], [7]). На рисунках в подобных случаях изображают только несколько первых шагов построения кривой. Что же до тех очевидных рассуждений, посредством которых можно доказывать теорему Жордана в различных частных случаях вроде окружности, квадрата и т. п., то эти рассуждения всякий раз оказываются тесно связанными со спецификой этих случаев, так что в различных случаях эти рассуждения совсем различны и не подсказывают, как быть в общем случае.

Но для читателя, который не особенно интересуется теоремой Жордана, а хотел бы узнать побольше о топологии, содержание моей книжки может быть интересно как простой пример небольшой топологической теории с её последующим использованием для доказательства утверждения, внешне не относящегося к этой теории. Говоря более общими словами, книжка в какой-то степени демонстрирует топологический стиль. Мне кажется, в этом отношении она даёт больше, чем [4] или упоминаемая ниже топологическая глава [5]. Что же до примеров топологических результатов, в [4] их больше. В связи с этим отмечу, что в [5] имеется небольшая глава о топологии, где на наглядном элементарном уровне рассказано о ещё бо́льшем числе вопросов. Не всё там доказывается, да и роль доказательств, в соответствии с общим характером книги, часто играют довольно наглядные рассуждения, интуитивно убедительные, но не всегда формально исчерпывающие. Наконец, значительно бо́льший в количественном отношении и качественно более глубокий материал содержится в [6]. Здесь тоже многое рассказывается, а не доказывается — полные доказательства потребовали бы развёртывания намного более сложного «технического аппарата», чем в настоящей книжке или в [4]. Но некоторое представление об общем характере этого аппарата авторы [6] попытались дать.

§ 2. Комплексные числа

Точки, радиус-векторы и координаты

Читателю должно быть известно, что «геометрическую» плоскость Π (с которой знакомятся в школьном курсе геометрии) можно «отождествить» с «арифметической плоскостью» $\mathbb{R}^2$, состоящей из пар (x, y) вещественных чисел [9]. Это делается с помощью декартовых координат: точке A плоскости сопоставляются её координаты, и тем самым получается соответствие между точками плоскости Π и парами чисел (x, y), причём каждой точке соответствует ровно одна числовая пара и каждая пара чисел отвечает ровно одной точке, что выражают словами: соответствие является взаимно однозначным. Условимся раз и навсегда выбрать в нашей плоскости Π какие-то декартовы координаты; тогда и соответствие между точками плоскости и числовыми парами будет установлено раз и навсегда.

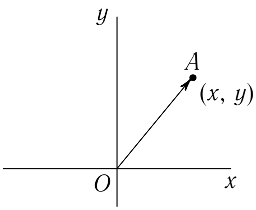


Рис. 5

Существенно, что при этом не просто устанавливается взаимно однозначное соответствие $\Pi \leftrightarrow \mathbb{R}^2$, но что важные геометрические понятия, относящиеся к Π, получают простую перефразировку в терминах, относящихся к парам чисел. В результате Π и $\mathbb{R}^2$ становятся

[9] $\mathbb{R}$ — это более или менее стандартное обозначение для множества вещественных чисел. Двойка в обозначении $\mathbb{R}^2$ указывает, что речь идёт о парах таких чисел. Рассматривая наборы $(x_1, x_2, \ldots, x_n)$ из n вещественных чисел, мы бы писали $\mathbb{R}^n$.

Раз уж речь зашла о парах и более длинных наборах, уместно заметить (или напомнить), что для любых двух множеств A, B через $A \times B$ обозначается их «прямое произведение» — множество пар (x, y), где $x \in A$, $y \in B$. В общем случае, для любых множеств $A_1, \ldots, A_n$ через $A_1 \times \ldots \times A_n$ обозначается множество наборов $(x_1, \ldots, x_n)$, где $x_1 \in A_1, \ldots, x_n \in A_n$. Обозначения $\mathbb{R}^2$ и $\mathbb{R}^n$ — это сокращения для $\mathbb{R} \times \mathbb{R}$ и $\underbrace{\mathbb{R} \times \ldots \times \mathbb{R}}_{n\ \text{раз}}$, аналогичные принятым в алгебре обозначениям для степеней.

как бы «моделями» друг для друга — говоря об одной из них, мы фактически говорим и о другой.

В полном объёме переход от «геометрии» (в Π) к арифметике и алгебре (в $\mathbb{R}^2$) и обратно составляет основную идею так называемой аналитической геометрии, изучаемой на первом курсе и затем используемой почти на каждом шагу как в математике, так и в большинстве её приложений. Нам из всех этих перефразировок нужны только две.

1. Мы говорим о точках плоскости Π, но раз мы пользуемся декартовыми координатами и, в частности, у нас выбрано начало координат O, то каждая точка A плоскости однозначно характеризуется своим радиус-вектором $\overrightarrow{OA}$. Координаты точки A — это также и координаты вектора $\overrightarrow{OA}$. Очень часто переходят от точек к радиус-векторам и обратно, не оговаривая этого особо, а просто подразумевая. Заметим, что при самом таком переходе мы, собственно, ещё не пользуемся координатными осями, а только используем начало координат O как своего рода «начало отсчёта» — исходную начальную точку для векторов. Соответствие между точками и их радиус-векторами зависит, конечно, от выбора O. Если выбрать другое начало отсчёта O', то точке A будет соответствовать вектор $\overrightarrow{O'A} = \overrightarrow{OA} + \overrightarrow{O'O}$, так что радиус-векторы всех точек изменятся на одно и то же векторное слагаемое [10].

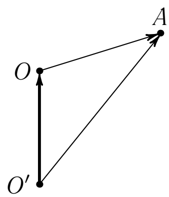


Рис. 6

Если быть педантичным, то можно считать, что у нас имеется «точечная» плоскость Π (состоящая, как подсказывает название, из точек) — в геометрии это наш исходный объект — и «векторная» плоскость Π_v (состоящая из векторов [11] плоскости Π). Выбрав начало отсчёта $O \in \Pi$, мы устанавливаем взаимно однозначное соответствие $\Pi \leftrightarrow \Pi_v$. Но если уж мы выбрали O, то точки и векторы становятся как бы в известной степени взаимозаменяемыми. Мы часто будем иметь в виду под элементами Π не столько точки, сколько векторы.

[10] Конечно, в геометрии и физике встречаются векторы, которые появляются, так сказать, сами по себе, независимо от радиус-векторов точек; они никак не зависят от выбора начала отсчёта. В геометрии таков вектор $\overrightarrow{AB}$ с данными началом A и концом B. В физике таковы вектор скорости или вектор напряжённости электрического поля. Но сейчас мы имеем дело с радиус-векторами.

[11] Речь идёт о «свободных» векторах, которые по известным правилам можно переносить из точки в точку. Встречаются величины векторной природы, которые по самому своему характеру связаны с определёнными «точками приложения». Наиболее известный пример такого рода относится к физике — это сила. (Сила в 1 кГ, приложенная к плечу взрослого человека, почти не заметна, но что, если её приложить к уху?) В подобных случаях говорят о «связанных векторах».

Как известно, существенное преимущество векторов перед точками состоит в том, что для векторов имеются две алгебраические операции: их можно складывать и умножать на обычные вещественные числа (или, как ещё говорят, на «скаляры»; скаляры — это просто числа, название «скаляры» обычно употребляют, когда хотят отметить их отличие от векторов) [12]. Для пар чисел тоже можно ввести аналогичные операции:

$$(x,\, y) + (u,\, v) = (x + u,\, y + v), \qquad a(x,\, y) = (ax,\, ay).$$

Известно, что при соответствии $\Pi \leftrightarrow \mathbb{R}^2$ указанные операции над векторами переходят в соответствующие операции над числовыми парами.

2. Расстояние $\rho(A,\, B)$ между точками A и B плоскости, имеющими координаты $(x,\, y)$ и $(u,\, v)$, выражается через эти координаты по формуле, легко получающейся с помощью теоремы Пифагора:

$$\rho(A,\, B) = \sqrt{(x-u)^2 + (y-v)^2}. \tag{1}$$

Можно принять, что в $\mathbb{R}^2$ правая часть (1) определяет «расстояние» между любыми двумя парами чисел $(x,\, y)$ и $(u,\, v)$. Тогда получается, что при соответствии $\Pi \leftrightarrow \mathbb{R}^2$ расстояние переходит в расстояние. Длина $|\boldsymbol{u}|$ вектора $\boldsymbol{u}$, имеющего координаты $(x,\, y)$ [13], выражается и того проще:

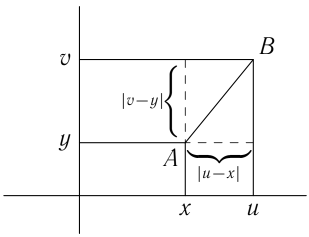


Рис. 7

$$|\boldsymbol{u}| = \sqrt{x^2 + y^2}.$$

Помимо декартовых координат на плоскости, довольно часто используют так называемые полярные координаты. Возьмём какую-нибудь точку O (на сей раз её называют полюсом) и проведём какой-нибудь луч (полупрямую) Ox, начинающийся в полюсе. Общепринятого названия для этого луча нет; я буду называть

[12] Мы можем временно принять, что это также и операции над точками, потому что мы условились раз и навсегда пользоваться одной и той же системой декартовых координат в Π и, в частности, одним и тем же началом отсчёта, так что слова «точка» и «вектор» стали практически равнозначными. Но в этом есть нечто искусственное; сейчас нас эта искусственность не тревожит, но вообще-то операции над точками зависели бы от выбора O, что не геометрично.

[13] Если бы мы должны были педантично различать точки и векторы, то это стоило бы делать и применительно к их координатам. Когда я начал учиться в университете, такое различие проводилось в курсе аналитической геометрии, где писали $(x,\, y)$ для координат точки и $\{x,\, y\}$ для координат вектора. Последовательно проводя подобный подход, мы могли бы сказать, что у нас есть два экземпляра арифметической плоскости $\mathbb{R}^2$ — «точечная» $\mathbb{R}^2_p$ (от point) и «векторная» $\mathbb{R}^2_v$. Но позднее на это уже не обращалось внимания, потому что различие либо не было существенным для имевшихся целей, либо было ясно из контекста.

его начальным лучом. Если мы уже работаем с некоторой декартовой системой координат в П, то естественно (хотя и не обязательно) принять начало координат за полюс, а положительную полуось абсцисс — за начальный луч. Если же мы имеем дело с арифметической плоскостью $\mathbb{R}^2$, то за O принимаем точку $(0, 0)$, а за начальный луч — луч, состоящий из точек $(x, 0)$, $x \geqslant 0$.

Положение любой точки A на плоскости можно охарактеризовать, указав её расстояние r до полюса O («полярное расстояние») и угол φ между начальным лучом Ox и лучом OA. Здесь надо договориться, как отсчитывать этот угол, который называют «полярным углом».

Условимся отсчитывать его в направлении против часовой стрелки. Такое направление отсчёта углов считается «положительным», а противоположное направление (т. е. направление по часовой стрелке) — «отрицательным». (Подразумевается обычное соглашение, что в П положительная полуось ординат получается из положительной полуоси абсцисс при повороте последней вокруг начала координат на прямой угол против часовой стрелки. Так оно и есть при обычном рисунке координатных осей, когда ось x горизонтальна и её положительное направление — это направление направо, а ось y вертикальна и её положительное направление — это направление вверх. А применительно к $\mathbb{R}^2$ положительным считается направление от начального луча к лучу, состоящему из точек $(0, y)$, $y \geqslant 0$. При нашем отождествлении $\Pi \leftrightarrow \mathbb{R}^2$ оно как раз и отвечает направлению против часовой стрелки в П.) Тогда, конечно, надо допустить, что угол может принимать значения от 0° до 360° или, если измерять угол в радианах (как мы и будем делать), от 0 до 2π. Пара (r, φ) — это и есть «полярные координаты» точки A; r называют полярным радиусом или радиальной координатой точки A, а φ — полярным углом или угловой координатой этой точки. В терминах полярных координат, $r = a$ — это уравнение окружности радиуса a с центром в полюсе, а $\varphi = \alpha$ (с постоянным α) — это уравнение некоторого луча с началом в O.

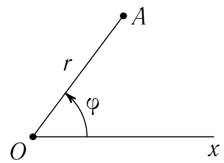


Рис. 8

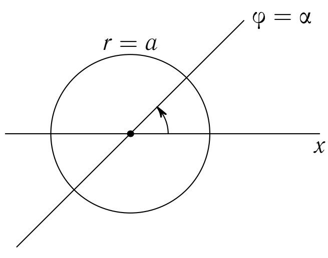


Рис. 9

Здесь есть ещё одна условность. Если точка A вращается вокруг O, скажем, в положительном направлении, т. е. против часовой стрелки, и пересекает начальный луч, то придётся признать, что при этом пересечении полярный угол скачком меняется от 2π до 0, а потом опять начинает возрастать. Иметь дело со скачком неприятно, и чтобы этого избежать, часто принимают, что при пересечении точкой A луча Ox полярная координата не делает скачка, а продолжает увеличиваться. А при вращении в противоположным направлении мы должны были бы считать, что полярный угол φ уменьшился до 0, скачком принял значение 2π и затем вновь убывает. Вместо этого удобно допустить для φ отрицательные значения. Поскольку A может сделать сколько угодно оборотов вокруг O, то приходится принять, что полярный угол φ может быть любым числом. Но, конечно, два угла, различающиеся на целочисленное кратное 2π, геометрически соответствуют одному и тому же направлению, т. е. одному и тому же лучу с началом в O. В связи с этим говорят, что φ «отсчитывается по модулю 2π» [14].

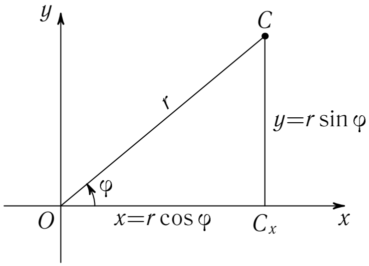


Рис. 10

Отметим простую связь между декартовыми и полярными координатами, которая имеет место при наших обычных соглашениях о расположении координатных осей, полюса и начального луча. Для точки C с декартовыми координатами (x, y) и полярными (r, φ) имеем

$$x = r\cos\varphi, \quad y = r\sin\varphi. \tag{2}$$

Когда C лежит в первом квадранте, в равенстве (2) легко убедиться, рассматривая прямоугольный треугольник OCC_x (где опять $CC_x \perp Ox$), в котором угол с вершиной O равен φ, $|OC| = r$ и $|OC_x| = x$, $|C_xC| = y$.

В других квадрантах получаются другие рисунки, x и y могут быть отрицательными, а угол $\angle COC_x = 180^\circ - \varphi$, $\varphi - 180^\circ$ или $360^\circ - \varphi$, так что надо вспомнить, как связаны друг с другом синусы и косинусы таких углов; однако окончательный результат тот же. (Что и не удивительно. Косинус и синус угла, бо́льшего 90°, определяются таким образом, чтобы у той точки единичной окружности с центром O, полярный угол которой

[14] Выражение «по модулю» применялось сперва в теории чисел, где говорили, что два целых числа k и l «сравнимы по модулю ненулевого целого числа n» и писали $k \equiv l \bmod n$, если разность $k - l$ делится (в классе целых чисел) на n, т. е. $k - l = mn$ с некоторым целым m. (Например, $17 \equiv 25 \bmod 4$.) Затем это выражение стали применять шире; в частности, для вещественных чисел a, b, c, где $c \neq 0$ говорят, что $a \equiv b \bmod c$, если $a - b = mc$ с некоторым целым m. Формулировка типа «φ отсчитывается по модулю 2π» означает, что вместе с φ ту же роль (в данном случае — роль угловой координаты некоторой точки) играют все числа вида $\varphi + 2\pi m$.

Как видно, слово «модуль» имеет два смысла (другой — то же, что и «абсолютная величина числа» или «длина вектора»). Они оба одинаково стандартны, так что приходится мириться со связанным с этим неудобством. Как уже было сказано по другому поводу, опыт показывает, что при достаточном внимании к контексту путаницы не возникает.

равен φ, декартовы координаты были $(\cos\varphi, \sin\varphi)$; возможно, читатель знаком именно с таким определением тригонометрических функций. А у точки, расположенной в r раз дальше от O (но тоже лежащей на луче с данным φ), x и y должны быть в r раз больше.)

Комплексная плоскость

Комплексные (или, как долго говорили, мнимые[15]) числа появились в середине XVI в. благодаря смелости Дж. Кардано и Р. Бомбелли, желавших, чтобы из отрицательных чисел можно было извлекать квадратный корень[16]. Однако они долго оставались чем-то таинственным. Так, Г. Лейбниц писал: «Мнимые числа — это прекрасное и чудесное убежище божественного духа, почти что амфибия[17] бытия с небытием». (Даром что к тому времени «амфибии» исполнилось полтора века.) Только около 1800 г. в природу «амфибии» была внесена ясность.

Напомню, что вещественные числа можно изображать точками прямой — скажем, оси абсцисс. Числу x сопоставляется точка этой прямой с координатой x (а если вспомнить, что наша прямая есть ось абсцисс на плоскости, то это будет точка с координатами $(x, 0)$). Для чисел вида $x+yi$ (где i — мнимая единица; $i^2 = -1$) на прямой нет места. Но для них есть место на плоскости: изобразим такое число точкой с координатами (x, y) или, действуя более формально, скажем, что число $x + yi$ (вернее сказать, то число, которое со временем мы сможем записать в виде $x + yi$) — это точка (x, y) «арифметической» плоскости $\mathbb{R}^2$.

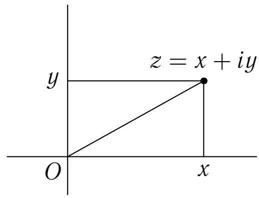


Рис. 11

В частности, теперь мнимая единица i — это единичный вектор оси ординат (направленный в сторону возрастания y), а более формально — пара чисел $(0, 1)$. Говорят, что x — это вещественная часть комплексного числа $z = x + yi$, а y — мнимая часть z; используют обозначения $x = \operatorname{Re} z$, $y = \operatorname{Im} z$. Сложение комплексных чисел и их умножение на ве-

[15] Теперь название «мнимые» (или «чисто мнимые») числа сохраняют только за комплексными числами вида ai с вещественными a.

[16] Вообще-то Кардано интересовался кубическими уравнениями (позднее к ним добавились и уравнения четвёртой степени), но по ходу дела появлялся некоторый квадратный корень, и оказалось, что иногда подкоренное выражение может быть отрицательным, а в то же время изучаемое кубическое уравнение имеет вещественные решения. Кардано решил, что всё-таки квадратный корень существует, только это какое-то особое (по выражению Кардано, «поистине софистическое») число, и что искомое вещественное решение кубического уравнения связано с этим числом так же, как если бы всё было вещественным. Сам Кардано не очень настаивал на этой идее, но Бомбелли понял её всерьёз.

[17] Здесь это слово использовано в старинном значении — как нечто промежуточного, двоякого характера.

щественные числа определяются как соответствующие операции для векторов или числовых пар. После этого сразу ясно, что

$$(x,\, y) = x(1,\, 0) + y(0,\, 1) = x \cdot 1 + yi = x + yi.$$

Пока что мы не имеем ничего нового сравнительно с векторной алгеброй. Новое связано с умножением комплексных чисел друг на друга (а не только на вещественные числа). Мы хотим, чтобы при умножении выполнялись обычные правила алгебры и было верно равенство $i^2 = -1$. Значит, мы хотим, чтобы было

$$(x + yi)(u + vi) = xu + yivi + yiu + xvi = \\ = xu + yvi^2 + (yu + xv)i = (xu - yv) + (yu + xv)i.$$

Вот мы и определим на нашей плоскости $\mathbb{R}^2$ умножение следующим образом:

$$(x,\, y)(u,\, v) = (xu - yv,\, yu + xv).$$

Когда в $\mathbb{R}^2$ введена такая новая операция, то говорят уже не о векторах или числовых парах, а о комплексных числах, а $\mathbb{R}^2$ с таким законом умножения обозначают через $\mathbb{C}$. Можно проверить, что умножение коммутативно (т. е. $zw = wz$) и ассоциативно (т. е. $(zw)t = z(wt)$; ассоциативность позволяет писать просто zwt, не заботясь о расстановке скобок), что оно связано со сложением согласно обычному правилу дистрибутивности (т. е. $z(w+t) = zw+zt$), что 1 (прежняя единица в $\mathbb{R}$, которая стала теперь точкой $(1,\, 0)$) является единицей во всём $\mathbb{C}$ (т. е. $1 \cdot z = z \cdot 1 = z$ для всех z), что $i^2 = -1$ и что для любого $z \neq 0$ имеется обратное число z^{-1} (для него $zz^{-1} = z^{-1}z = 1$); имея z^{-1}, определяем деление на z (обозначаемое, как обычно, с помощью черты или двоеточия) как умножение на z^{-1}. Всё это требует места и времени, но делается более или менее автоматически. Я на этом останавливаться не буду.

А вот на чём я хочу на момент остановиться — это на различии между комплексными числами и векторами. Я сказал, в чём это различие состоит: комплексные числа мы можем умножать друг на друга, а векторы — нет. Но почему бы не определить умножение векторов как умножение соответствующих комплексных чисел — разве это не обогатило бы векторную алгебру?

Беда в том, что такое определение с самого начала зависит от выбранной системы координат на плоскости. А «идеология» векторной алгебры состоит в том, что векторные операции должны определяться независимо от того, какие именно координаты используются. Правда, возникает вопрос: а действительно ли при определении умножения векторов как умножения соответствующих комплексных чисел результат будет зависеть от используемых координат? Ведь в правой части формулы (1) фигурируют координаты, а результат — расстояние между двумя точками — от выбора координат не зависит.

Упражнение 1. При перемножении двух вещественных чисел, т. е. точек оси абсцисс, снова получается вещественное число, т. е. точка оси абсцисс. Покажите, что нет никакой другой прямой, которая обладала бы аналогичным свойством «сохранения при умножении на себя»: произведение любых её точек снова лежит на этой прямой.

Получается, что при использовании нашего умножения осью абсцисс должна быть именно такая-то прямая и никакая другая. А если мы повернём систему координат и вновь определим умножение, пользуясь новыми координатами, то это будет уже другое умножение, при нём «сохраняться при умножении на себя» будет другая прямая (а именно, новая ось абсцисс).

Геометрическая трактовка комплексных чисел быстро завоевала признание благодаря авторитету опубликовавшего её «короля математиков» К. Гаусса. (Однако ещё до него её предложили по крайней мере трое. Первым эту идею высказал английский математик Дж. Валлис, старший современник И. Ньютона. Почему-то на неё тогда не обратили внимания, хотя Валлис был заметной фигурой; может быть, причина в том, что сам он не стал подробнее развивать свою идею. Позднее по тому же пути независимо последовали два малоизвестных математика-любителя — сперва К. Вессель, работа которого была благополучно забыта на 100 лет, затем Ж. Арган, которому повезло больше — на его работу обратили внимание всего через 7 лет после публикации, когда геометрический подход к $\mathbb{C}$ всё ещё был новым.) Замечу, что всё же для комплексных чисел можно дать совсем другие исходные определения[18]; тогда наши построения окажутся не определениями (которые уже будут иметься в наличии), а очень полезной интерпретацией комплексных чисел. Польза её в том, что мы сразу наглядно представляем себе всю систему комплексных чисел $\mathbb{C}$ и сразу видим, что при работе с последними можно использовать геометрию плоскости. Обратно, алгебра комплексных чисел может помочь при исследовании геометрии плоскости (мы как раз и будем этим пользоваться).

В частности, так называемое «абсолютное значение» или «норма» $|z|$ комплексного числа $z = x + yi$ — это просто длина его как вектора (т. е. $\sqrt{x^2 + y^2}$). Расстояние между комплексными числами z и w — это

[18] О. Коши, с именем которого связаны (помимо всего прочего) большие достижения по части «наведения порядка» в математическом анализе и теории функций комплексного переменного, придерживался (по крайней мере, в части своих работ) не геометрического определения комплексных чисел, а чисто алгебраического, которое в современных терминах гласит: поле комплексных чисел $\mathbb{R}$ есть факторкольцо $\mathbb{R}[x]/\{x^2 + 1\}$ кольца $\mathbb{R}[x]$ всех многочленов с вещественными коэффициентами по идеалу $\{x^2 + 1\}$, порождённому многочленом $x^2 + 1$.

норма $|z - w|$ их разности (чаще прямо о ней и говорят). При отражении в прямой $\mathbb{R}$ (вещественные числа рассматриваются как частный случай комплексных; $\mathbb{R}$ — это ось абсцисс на комплексной плоскости $\mathbb{C}$), при котором каждая точка переходит в симметричную ей относительно этой прямой точку, число $z = x + yi$ переходит в «комплексно сопряжённое» число $\overline{z} = x - yi$. Заметим, что $|z|^2 = z\overline{z}$.

Когда мы записываем комплексное число z в виде $x + yi$, это отвечает использованию декартовых координат на плоскости $\mathbb{C}$. А использованию полярных координат (с полюсом в нуле и вещественной положительной полуосью в качестве начального луча) отвечает представление z в тригонометрической форме

$$z = r(\cos\varphi + \sin\varphi), \quad \text{где} \quad r = |z|.$$

(r, φ) — это полярные координаты точки z; их связь с декартовыми координатами на плоскости выражается по формулой (2). Отметим ещё раз, что точки вида $\cos\varphi + i\sin\varphi$ суть точки единичной окружности (см. рис. 12),

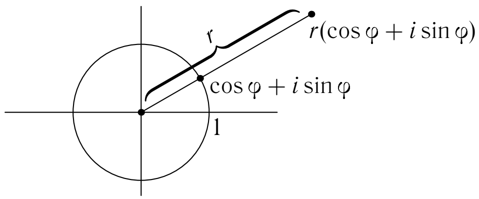


Рис. 12

т. е. окружности единичного радиуса с центром в нуле (раньше то же самое было выражено иными словами — говорилось о точках с декартовыми координатами $(\cos\varphi, \sin\varphi)$). В порядке стандартизации мы будем обозначать[19] эту окружность через $\mathbb{S}^1$.

Имея дело с комплексными числами, φ называют не полярным углом, а аргументом комплексного числа z. Будучи по существу полярным углом, аргумент определён не однозначно, а только с точностью до целочисленного кратного 2π; таким образом, аргументов у z много. В порядке стандартизации принято обозначать через $\arg z$ значение φ, для которого $0 \leqslant \varphi < 2\pi$. Однако в этой брошюре нас больше устроило бы $-\pi < \varphi \leqslant \pi$. Но если пользоваться стандартным обозначением $\arg z$, то приходится следовать и стандартному соглашению о значениях этого $\arg z$. Поэтому я буду

[19] Тогда как произвольная топологическая окружность у нас чаще всего будет обозначаться через Σ^1, но это обозначение, хотя оно и довольно естественное, отнюдь не является стандартным.

обозначать «наш» аргумент через $\widehat{\arg} z$ (обозначение не является стандартным); для него, повторяю, $-\pi < \widehat{\arg} z \leqslant \pi$ (см. рис. 13). Для $z = 0$ аргумент не определён.

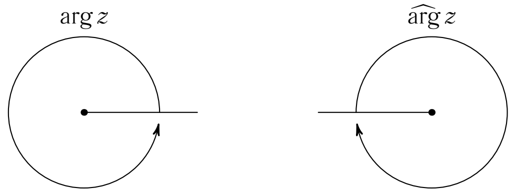


Рис. 13

Упражнение 2. Докажите, что если z не находится на отрицательной полуоси (включая в последнюю и 0), то $\widehat{\arg} z = \arg(-z) - \pi$. (Указание. Пусть $\varphi = \widehat{\arg} z$; при этом $-\pi < \varphi < \pi$ (равенство $\varphi = \pi$ мы исключили, оговорив, что $z \notin (-\infty, 0]$). Тогда $\varphi + \pi$ — один из аргументов числа $-z$. В каком интервале заключено число $\varphi + \pi$?)

Функция $\arg z$ от переменной z имеет разрывы в точках положительной вещественной полуоси (включая в последнюю и 0), а $\widehat{\arg} z$ — в точках отрицательной вещественной полуоси (тоже включая в неё 0). В остальных точках функции $\arg z$ и $\widehat{\arg} z$ непрерывны. Непрерывность $\arg z$ в точках $\mathbb{C} \setminus [0, \infty)$ следует из того очевидного геометрического факта, что маленький кружочек с центром z_0 находится внутри малого угла с вершиной в 0 [20]. (При желании на базе обычного курса математического анализа можно дать и аналитическое доказательство непрерывности, обращающееся к геометрии только в той степени, в какой это нужно для определения $\arg z$, но мы этого делать не будем.) Непрерывность $\widehat{\arg} z$ в точках $\mathbb{C} \setminus (-\infty, 0]$ можно доказать аналогично, а можно и вывести её из непрерывности $\arg z$, используя упражнение 2.

Повторяю ещё раз (теперь уже на «комплексном» языке), что при описании движения в $\mathbb{C} \setminus \{0\}$ точки $z(t)$, непрерывно зависящей от t, целесообразно выбрать её аргумент $\varphi(t)$ таким образом, чтобы он непрерывно зависел от t. Впоследствии мы убедимся, что это возможно, но при этом соглашение, согласно которому $0 \leqslant \varphi < 2\pi$, может не выполняться.

---

[20] Пусть $z_0 \notin [0, \infty)$, $\varphi = \arg z_0$ (так что $0 < \varphi < 2\pi$). Докажем непрерывность функции $\arg z$ в точке $z = z_0$. Для заданного $\varepsilon > 0$ мы должны указать такое $\delta > 0$, что если $|z - z_0| < \delta$, то $|\arg z - \varphi| < \varepsilon$. Уменьшив, если понадобится, ε, мы вправе считать, что $0 < \varphi - \varepsilon$ и $\varphi + \varepsilon < 2\pi$. Существует такое $\delta > 0$, что если $|z - z_0| < \delta$ (т. е. если z лежит в круге радиуса δ с центром в z_0), то угол α между лучами $0z_0$ и $0z$ меньше ε. Геометрически очевидно также, что луч $0z$ образует с положительной полуосью x угол $\varphi \pm \alpha$, где знак зависит от того, получается ли луч $0z$ из луча $0z_0$ при повороте на угол α в положительном или отрицательном направлении (т. е. против часовой стрелки или по ней). Значит, $\varphi \pm \alpha$ — это один из аргументов числа z. Но $0 < \varphi \pm \alpha < 2\pi$; значит, $\arg z = \varphi \pm \alpha$, $|\arg z - \varphi| = \alpha < \varepsilon$.

Тригонометрическая форма комплексных чисел удобна при умножении, потому что[21)]

$$r_1(\cos\varphi_1+i\sin\varphi_1)\cdot r_2(\cos\varphi_2+i\sin\varphi_2) = r_1r_2(\cos(\varphi_1+\varphi_2)+i\sin(\varphi_1+\varphi_2)) \quad (3)$$

(см. рис. 14). В частности, $i=\cos\dfrac{\pi}{2}+i\sin\dfrac{\pi}{2}$, поэтому при умножении z на i аргумент числа z увеличивается на $\dfrac{\pi}{2}$, норма же (т. е. абсолютная величина) не меняется. Иными словами, умножение на i с геометрической

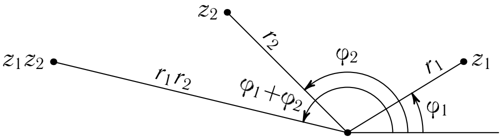


Рис. 14. Умножение z_1z_2

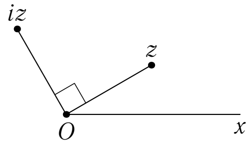


Рис. 15

точки зрения состоит в повороте вокруг 0 на прямой угол в положительном направлении (см. рис. 15).

Некоторые понятия математического анализа в комплексной области

Поскольку у нас имеется понятие расстояния между двумя комплексными числами, в комплексной области можно пользоваться понятиями, связанными с непрерывностью, пределами и т. п. Так, последовательность z_n сходится к z, а z является пределом z_n, если $|z-z_n|\to 0$ при $n\to\infty$, т. е. если для любого $\varepsilon>0$[22)] существует такое N, что $|z-z_n|<\varepsilon$ при всех $n>N$. Это равносильно сходимости двух вещественных последовательностей $\operatorname{Re} z_n$ и $\operatorname{Im} z_n$ к $\operatorname{Re} z$ и $\operatorname{Im} z$. Вот необходимое и достаточное условие для сходимости (критерий Коши), дословно совпадающее с условием для вещественного случая[23)]: для любого $\varepsilon>0$ существует такое N, что $|z_n-z_m|<\varepsilon$ при любых n, $m\geqslant N$. (Другими словами, члены последовательности с достаточно большими номерами очень близки друг к другу.) Можно говорить и о пределе $\lim\limits_{z\to z_0} f(z)$ функции $f(z)$ от комплексного пере-

[21)] Раскрывая скобки в левой части, можно выразить вещественную и мнимую части произведения через тригонометрические функции от φ_1, φ_2. Вычисление показывает, что вещественная часть совпадает с известным выражением для косинуса суммы, а мнимая — для её синуса.

[22)] Написав, что $\varepsilon>0$, мы сказали тем самым также и то, что число ε — вещественное, ибо только для вещественных чисел и определено понятие «больше».

[23)] Доказательство легко свести к вещественному случаю, рассматривая две последовательности, образованные вещественными и мнимыми частями $\operatorname{Re} z_n$, $\operatorname{Im} z_n$ комплексных чисел z_n.

менного z, определённой на некотором подмножестве M плоскости $\mathbb{C}$. Этот предел есть такое число a, что при достаточной близости z к z_0 разность $f(z) - a$ становится сколь угодно малой по абсолютной величине (т. е. $f(z)$ подходит сколь угодно близко к a). Подробнее это означает следующее: для любого $\varepsilon > 0$ существует такое $\delta > 0$, что когда $z \in M$ и $|z - z_0| < \delta$, то $|f(z) - a| < \varepsilon$. При этом говорят также, что $f(z)$ стремится к a (запись: $f(z) \to a$) при $z \to z_0$. Функция $f(z)$ (по-прежнему определённая на M) называется непрерывной в точке $z_0 \in M$, если $\lim\limits_{z \to z_0} f(z) = f(z_0)$. Когда f непрерывна во всех точках M, то просто говорят, что f — непрерывная функция.

Далее, если для всех целых $n \geqslant k$ заданы комплексные числа a_n, то можно говорить о «ряде» или «бесконечной сумме»

$$\sum_{n=k}^{\infty} a_n = a_k + a_{k+1} + a_{k+2} + \ldots + a_n + \ldots$$

Говорят, что этот ряд сходится к числу a, если частичные суммы этого ряда, — т. е. конечные суммы

$$s_m = \sum_{n=k}^{m} a_n = a_k + a_{k+1} + a_{k+2} + \ldots + a_m,$$

— стремятся к a при $m \to \infty$. Говорят, что a есть сумма данного ряда, и пишут $\sum\limits_{n=k}^{\infty} a_n = a$. Как видно, понятие сходимости и суммы ряда — просто своего рода перефразировка понятий сходимости последовательности и её предела. Новым является понятие абсолютной сходимости, относящееся уже специально к рядам: ряд $\sum\limits_n a_n$ сходится абсолютно, если сходится ряд $\sum\limits_n |a_n|$. В этом случае сходится и исходный ряд $\sum\limits_n a_n$ (почему?). (Примените критерий сходимости Коши к последовательности его частичных сумм.) Абсолютно сходящиеся ряды во многих отношениях похожи на обычные конечные суммы (суммы конечного числа слагаемых), тогда как свойства тех рядов, которые сходятся не абсолютно (их называют сходящимися условно), бывают довольно «капризными». Нам не придётся переживать из-за этих «капризов», так как те немногие ряды, с которыми мы будем иметь дело, сходятся абсолютно. Мы будем иметь дело только со степенными рядами, члены которых суть различные степени комплексного переменного z с некоторыми множителями, т. е. рядами вида $\sum\limits_n a_n z^n$. (Степенными рядами называют также ряды чуть более общего вида $\sum\limits_n a_n (z - z_0)^n$.) В математическом анализе устанавливают несколько важных специальных свойств степенных рядов, но мы не будем ссылаться на эти общие результаты: в нужном нам случае мы непосредственно получим то немногое, что нам потребуется.

Наконец, важнейшим понятием анализа является производная. Её определение дословно переносится на комплексный случай: производная функции $f(z)$ комплексного переменного z в точке z_0 есть

$$f'(z_0) = \frac{df}{dz}(z_0) = \lim_{h\to 0} \frac{f(z_0+h) - f(z_0)}{h}.$$

(Подразумевается, что $f(z)$ определена для всех z, достаточно близких к z_0.) Если функция имеет производную в точке z_0, то говорят, что эта функция дифференцируема в данной точке. Конечно, теперь мы не можем наглядно представлять себе z как время, $f(z)$ — как величину, изменяющуюся со временем, а $f'(z_0)$ — как мгновенную скорость изменения данной величины. Но другое наглядное представление, относящееся к производной, остаётся в силе: в маленькой области возле z_0 функция f является почти линейной, она почти совпадает с функцией $f(z_0)+f'(z_0)(z-z_0)$. Говоря точнее, $f'(z_0)h$ является главной линейной частью приращения функции $f(z)$ возле z_0, т. е. $f(z_0+h) = f(z_0) + f'(z_0)h + r(h)$, где остаток $r(h)$ (от английского «rest») имеет по h порядок малости выше первого. (Последнее, как обычно, означает, что $\lim\limits_{h\to 0} r(h)/h = 0$.)

Подобного рода понятия и связанные с ними факты совершенно аналогичны понятиям и фактам, относящимся к вещественной области и, вероятно, в какой-то степени известны читателю. (Это не означает, что «комплексный» математический анализ вообще аналогичен «вещественному». Нет, с некоторого места в комплексном анализе начинаются настоящие чудеса по сравнению с вещественным. Просто мы находимся в самой начальной части комплексного анализа, где никаких чудес ещё нет.)

Экспонента в комплексной области

В вещественной области известно разложение экспоненциальной функции e^x в ряд:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \ldots + \frac{x^n}{n!} + \ldots \qquad (x \in \mathbb{R}), \tag{4}$$

который абсолютно сходится при всех x. (Если читателю оно ещё не известно, не беда — мы ссылаемся на него только в порядке наводящего соображения; если этого ряда не знать, наши действия будут выглядеть менее мотивированными, вот и всё.) Этот ряд сходится при всех $x \in \mathbb{R}$. Л. Эйлер предложил определить e^z (пишут также $\exp z$) для комплексных z

посредством такой же формулы:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \frac{z^2}{2} + \frac{z^3}{6} + \ldots + \frac{z^n}{n!} + \ldots \qquad (z \in \mathbb{C}). \tag{5}$$

Упражнение 3. Докажите абсолютную сходимость ряда (5) при всех z. (Указание. Пусть имеются такие два ряда $\sum\limits_{n=k}^{\infty} a_n$ и $\sum\limits_{n=k}^{\infty} b_n$, что все $|b_n| \leqslant a_n$ (чем сказано также, что числа a_n — вещественные неотрицательные); тогда говорят, что первый ряд мажорирует второй (а второй — мажорируется первым). Если при этом первый ряд сходится, то и второй тоже сходится, и притом абсолютно. (Сравните $\left|\sum\limits_{n=k}^{m} |b_n| - \sum\limits_{n=k}^{l} |b_n|\right|$ с аналогичной разностью для первого ряда.) Условие, что $|b_n| \leqslant a_n$ при всех n, можно чуть ослабить: $|b_n| \leqslant a_n$ при всех достаточно больших n (т. е. при всех n, бóльших некоторого N). Используя это замечание, сравните ряд (5) с подходящим заведомо сходящимся рядом, члены которого положительны, скажем, с рядом для $e^{|z|}$ (сходимость которого, возможно, вам уже известна) или с $\sum\limits_{n=0}^{\infty} \frac{1}{2^n}$ (тогда можно ничего не знать о ряде (4)).)

Важнейшее свойство экспоненты состоит в том, что (совершенно аналогично соответствующему свойству в вещественной области)

$$e^z e^w = e^{z+w} \quad \text{при всех} \;\; z,\, w \in \mathbb{C}. \tag{6}$$

Чтобы доказать (6), будем для начала обращаться с рядами так же беззаботно, как если бы это были многочлены от z и w. Тогда можно написать

$$e^z e^w = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} = \sum_{n,m} \frac{z^n w^m}{n!\, m!}.$$

Сгруппируем здесь вместе члены с одинаковыми суммами показателей $n + m$. Обозначив такую сумму через k, видим, что k пробегает значения $k = 0, 1, 2, \ldots$ и что для каждого k имеется слагаемое с $n = 0$, $m = k$; $n = 1$, $m = n - 1$; $\ldots$; $n = k$, $m = 0$, словом, с $n = 0, 1, \ldots, k$, причём соответствующие m суть $k - n$. Итак,

$$e^z e^w = \sum_{k=0}^{\infty} \sum_{n=0}^{k} \frac{z^n w^{k-n}}{n!\,(k-n)!}.$$

Знаменатель наводит на мысль о числе сочетаний[24] $C_k^n = \dfrac{k!}{n!\,(k-n)!}$.

Умножим и одновременно разделим $\dfrac{z^n w^{k-n}}{n!\,(k-n)!}$ на $k!$; делитель $k!$ — один

[24] Это число также обозначают через $\binom{k}{n}$.

и тот же для всех слагаемых с данным k, и его можно вынести из-под знака суммирования по n, так что

$$e^z e^w = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{n=0}^{k} C_k^n z^n w^{k-n}.$$

Но $\sum\limits_{n=0}^{k} C_k^n z^n w^{k-n} = (z+w)^k$ (бином Ньютона), значит,

$$e^z e^w = \sum_{k=0}^{\infty} \frac{(z+w)^k}{k!} = e^{z+w},$$

что и требовалось доказать, — если только мы оправдаем наши беззаботные действия с бесконечными рядами.

В университетском курсе математического анализа даётся такое оправдание в довольно общем виде (однако всё же это не индульгенция на все возможные случаи перемножения рядов!). Но я не предполагаю известной эту общую теорему. Нам нужно доказать одно-единственное равенство

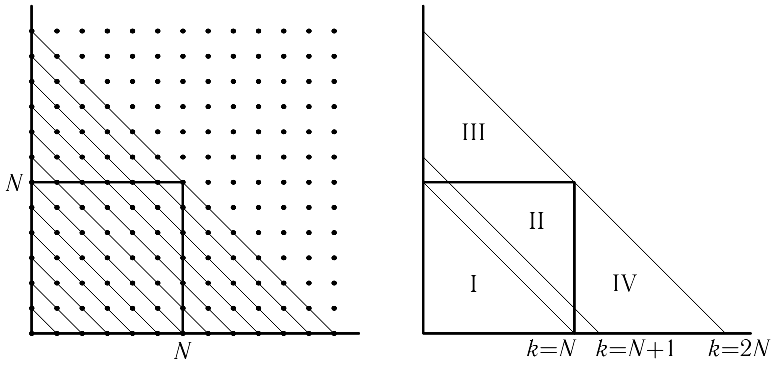
$$\sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{m=0}^{\infty} \frac{w^m}{m!} = \sum_{k=0}^{\infty} \frac{(z+w)^k}{k!}, \tag{7}$$

и мы это сделаем непосредственно, используя специфику интересующего нас ряда (5).

Обозначим частичную сумму ряда (5), получающуюся, когда берутся только слагаемые с $n = 0, 1, \ldots, N$, через $s_N(z)$. Это — многочлен от z, а $s_N(w)$ — аналогичный многочлен от w; перемножив их, мы вправе сгруппировать слагаемые по своему усмотрению. Но сперва посмотрим на произведение $s_N(z)s_N(w)$, не делая никакой группировки слагаемых:

$$s_N(z)s_N(w) = \sum_{0 \leqslant n,m \leqslant N} \frac{z^n w^m}{n!\,m!}.$$

Степени n, m, в которые возводятся здесь z и w — это натуральные числа, принимающие в нашей сумме значения от 0 до N. На рис. 16, а) точки изображают всевозможные пары натуральных чисел (n, m) и выделен квадрат, где $0 \leqslant m, n \leqslant N$. Теперь, аналогично тому как это (незаконно) делалось для бесконечных рядов, мы сгруппируем слагаемые с одинаковыми суммами показателей $n + m = k$. Здесь $0 \leqslant n, m \leqslant N$; значит, получатся суммы показателей $k = 0, 1, \ldots, 2N$. На рисунке пары (n, m) с $n + m = k$ при каждом k расположены на наклонной линии, параллельной одной из диагоналей квадрата. Некоторые из них не выходят за пределы квадрата, но у нас есть только слагаемые $\dfrac{z^n w^m}{n!\,m!}$ с такими (n, m), которые лежат в квад-



а) б)

Рис. 16

рате. Иными словами, при данном k показатель n у z^n уже не всегда будет пробегать все значения от 0 до k, — ведь $n \leqslant N$. Все значения $0, 1, \ldots, k$ показатель n пробегает, только когда $k \leqslant N$. Поэтому

$$s_N(z)s_N(w) = \sum_{k=0}^{N}\sum_{n=0}^{k}\frac{z^n w^{k-n}}{n!\,(k-n)!} + \sum_{k=N+1}^{2N}\sum_{\substack{0\leqslant n,m\leqslant N\\ n+m=k}}\frac{z^n w^{k-n}}{n!\,(k-n)!}.$$

В первой (двойной) сумме в правой части фигурирует показатель $(n,\ k-n)$ из треугольника I, условно помеченного знаком I на рис. 16, б), во второй — из треугольника II. Первую можно преобразовать так же, как это делалось выше, — только теперь это законно, ибо речь идёт о конечных суммах, — в $\sum\limits_{k=0}^{N}\dfrac{(z+w)^k}{k!} = s_N(z+w)$. Следовательно,

$$s_N(z)s_N(w) - s_N(z+w) = \sum_{k=N+1}^{2N}\sum_{\substack{0\leqslant n,m\leqslant N\\ n+m=k}}\frac{z^n w^{k-n}}{n!\,(k-n)!} =$$

$$= \sum_{k=N+1}^{2N}\frac{1}{k!}\sum_{\substack{0\leqslant n,m\leqslant N\\ n+m=k}} C_k^n z^n w^{k-n}. \qquad (8)$$

Разность $s_N(z)s_N(w) - s_N(z+w)$ стремится к $e^z e^w - e^{z+w}$ при $N \to \infty$, и если мы хотим доказать, что последняя разность равна нулю, то надо доказать, что правая часть (8) стремится к нулю при $N \to \infty$. Её норма не превосходит

$$\sum_{k=N+1}^{2N} \frac{1}{k!} \sum_{\substack{0\leqslant n,m\leqslant N\\ n+m=k}} C_k^n |z|^n |w|^{k-n}. \tag{9}$$

Показатели $(n,\, k-n)$ в (9) берутся из треугольника II. Фигурирующие здесь слагаемые — это часть слагаемых в

$$\sum_{k=N+1}^{2N} \frac{1}{k!} \sum_{n=0}^{k} C_k^n |z|^n |w|^{k-n}, \tag{10}$$

а если мы добавим к (9) новые положительные слагаемые, то сумма только увеличится. Вот мы и заменим (9) на (10). (Геометрически, мы добавили положительные слагаемые с показателями из III и IV.) А в последнем выражении сумма по n — это разложение $(|z|+|w|)^k$ в сумму одночленов по формуле бинома Ньютона. Итак,

$$|s_N(z)s_N(w) - s_N(z+w)| \leqslant \sum_{k=N+1}^{2N} \frac{(|z|+|w|)^k}{k!}.$$

Мы ещё более увеличим правую часть, если заменим $\sum\limits_{k=N+1}^{2N}$ на $\sum\limits_{k=N+1}^{\infty}$. Но тогда у нас получится «остаток» ряда для $e^{|z|+|w|}$:

$$\sum_{k=N+1}^{\infty} \frac{(|z|+|w|)^k}{k!} = \sum_{k=0}^{\infty} \frac{(|z|+|w|)^k}{k!} - \sum_{k=0}^{N} \frac{(|z|+|w|)^k}{k!} = \\ = e^{|z|+|w|} - s_N(|z|+|w|),$$

а о последнем выражении мы знаем, что оно стремится к 0 при $N \to \infty$. Тем самым мы доказали равенство (7).

Докажем непрерывность экспоненциальной функции. Так как $e^{z_0+h} - e^{z_0} = e^{z_0}(e^h - 1)$, то нужно только доказать, что $\lim\limits_{h\to 0} e^h = 1$. (Таким образом, (6) позволяет нам ограничиться проверкой непрерывности экспоненты в нуле.) А это совсем просто: ведь при $|h| < 1$

$$|e^h - 1| = \left|\sum_{n=1}^{\infty} \frac{1}{n!} h^n\right| \leqslant \sum_{n=1}^{\infty} \frac{1}{n!} |h^n| < \sum_{n=1}^{\infty} \frac{1}{n!} |h| = (e-1)|h|.$$

Упражнение 4. Докажите, что функция e^z дифференцируема во всех точках $\mathbb{C}$ и что её производная $(e^z)' = e^z$.

Формула Эйлера

Рассмотрим $e^{iz} = \sum\limits_{n=0}^{\infty} \frac{1}{n!}(iz)^n$. Здесь $(iz)^n = i^n z^n$, а степени числа i таковы:

$$i^0 = 1, \quad i^1 = i, \quad i^2 = -1, \quad i^3 = -i,$$
$$i^4 = 1, \quad i^5 = i, \quad i^6 = -1, \quad i^7 = -i,$$

и т. д. Когда n чётное, $n = 2j$ ($j = 0, 1, 2, \ldots$), то i^n — это вещественное число $(-1)^j$. Когда n нечётное, $n = 2j+1$ ($j = 0, 1, 2, \ldots$), то i^n — это мнимое число $(-1)^j i$. Разобьём ряд $\sum\limits_{n=0}^{\infty} \frac{1}{k!}(iz)^n$ в сумму двух рядов, содержащих слагаемые нашего ряда с чётными и нечётными n:

$$\sum_{n=0}^{\infty} \frac{1}{k!}(iz)^n = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!} z^{2j} + i \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} z^{2j+1}.$$

Эйлер знал, что при вещественных z стоящие справа ряды суть ряды для косинуса и синуса[25]:

$$\cos x = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!} x^{2j}, \quad \sin x = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} x^{2j+1} \qquad (x \in \mathbb{R}). \tag{11}$$

Поэтому он предложил определить косинус и синус для комплексных значений аргумента как суммы таких же рядов:

$$\cos z = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j)!} z^{2j}, \quad \sin z = \sum_{j=0}^{\infty} \frac{(-1)^j}{(2j+1)!} z^{2j+1} \qquad (z \in \mathbb{C}). \tag{12}$$

Эти ряды действительно сходятся при всех z и притом сходятся абсолютно (почему?). (Если читатель знаком с равенствами (11), то ему ясно, что таким образом Эйлер осуществил продолжение обычных тригонометрических функций в комплексную область. Если читатель с этим не знаком, то пока что он может считать, что Эйлер определил некие две функции в комплексной области, которые он почему-то обозначил так же, как обозначаются привычные тригонометрические функции. Некоторое время мы будем иметь дело именно с этими новыми функциями. Позднее мы объясним, как, не пользуясь рядами (11), доказать их совпадение с обычными косинусом и синусом для вещественных z.) Тогда получается формула Эйлера — одна их самых замечательных формул во всей математике:

$$e^{iz} = \cos z + i \sin z \qquad (z \in \mathbb{C}). \tag{13}$$

[25] Эйлер знал, но читатель, если он школьник, может и не знать. Прошу пока что принять это на веру.

Поскольку в ряд для косинуса входят только чётные степени z, а в ряд для синуса — только нечётные, то

$$\cos(-z) = \cos z, \quad \sin(-z) = -\sin z \qquad (z \in \mathbb{C}), \tag{14}$$

что выражают словами: косинус — чётная функция, а синус — нечётная. Поэтому из (13) следует, что $e^{-iz} = \cos z - i \sin z$, а это вместе с (13) приводит к такой перефразировке определения косинуса и синуса в комплексной области:

$$\cos z = \frac{1}{2}(e^{iz} + e^{-iz}), \quad \sin z = \frac{1}{2i}(e^{iz} - e^{-iz}).$$

Эти формулы не только следуют из сказанного, но и могут быть приняты за (внешне новое) определение косинуса и синуса, которое эквивалентно (12) (проверьте). Заметим, что теперь очевидна непрерывность и дифференцируемость косинуса и синуса во всех точках $\mathbb{C}$, ибо для экспоненты эти свойства уже установлены.

Упражнение 5. Докажите, что косинус и синус дифференцируемы всюду в $\mathbb{C}$, и что $(\cos x)' = -\sin x$, $(\sin x)' = \cos x$.

Упражнение 6. Докажите, что в комплексной области остаются в силе известные формулы тригонометрии

$$\begin{gathered} \cos^2 z + \sin^2 z = 1, \\ \cos(z + w) = \cos z \cos w - \sin z \sin w, \\ \sin(z + w) = \sin z \cos w + \cos z \sin w. \end{gathered} \tag{15}$$

(Указание к первой формуле. В вещественное области выражение $a^2 + b^2$ не разлагается на множители. А в комплексной?)

Формула Эйлера (13) верна при всех комплексных z. Остановимся особо на том частном случае, когда $z = t$ вещественно. При вещественных аргументах косинус и синус вещественны (если определять их согласно (12), то вещественность суммы ряда следует из вещественности всех его коэффициентов). Поэтому $\cos t$ — это вещественная часть e^{it}, а $\sin t$ — мнимая:

$$\cos t = \operatorname{Re} e^{it}, \quad \sin t = \operatorname{Im} e^{it} \tag{16}$$

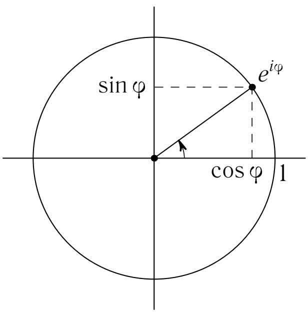


Рис. 17

(см. рис. 17). При вещественном t точка e^{it} лежит в плоскости $\mathbb{C}$ на единичной окружности $\mathbb{S}^1$ (ведь $|e^{it}|^2 = \cos^2 t + \sin^2 t = 1$). Если мы согласны, что при вещественном t эйлеровские $\cos t$ и $\sin t$ совпадают с обычными, то получается, что t — это аргумент (один из аргументов) комплексного числа e^{it}. Но теперь мы достигли такого уровня, что, и не обращаясь к (11), можем доказать все эти совпадения (совпадение t с одним из аргументов e^{it} и совпадение в вещественной области эйлеровских тригонометрических функций с обычными).

Представляя себе вещественное число t как время, рассмотрим движение точки e^{it} с изменением времени. Мы знаем, что движение происходит по единичной окружности. Его скорость

$$\frac{de^{it}}{dt} = \frac{de^z}{dt}\Big|_{z=e^{it}} \frac{d(it)}{dt} = ie^z. \tag{17}$$

(Здесь мы используем формулу для производной сложной функции:

$$\frac{d(f(z))}{dt} = \frac{df(z)}{dz}\Big|_{z=z(t)} \frac{dz(t)}{dt},$$

которая, вероятно, известна читателю в вещественном случае; в комплексном случае её доказательство такое же. Если, как в (17), $z(t)$ является линейной функцией от t, доказательство совсем просто.) По своей абсолютной величине скорость всё время является единичной. Движение всё время направлено против часовой стрелки, потому что вектор скорости ie^{it} получается при повороте радиус-вектора e^{it} на прямой угол против часовой стрелки. При $t = 0$ движущаяся точка находится в 1. Значит, за время t она «заметает» дугу единичной окружности длины t. Пока точка, сделав полный оборот по $\mathbb{S}^1$, не вернётся в 1, эта дуга начинается в точке 1 (где полярный угол равен 0) и кончается в e^{it}, причём направление вдоль дуги от её первого конца ко второму — положительное. Именно такой дугой и измеряется (в радианах) полярный угол точки e^{it}, т. е. аргумент $\arg e^{it}$ последнего комплексного числа. Мы видим, что $t = \arg e^{it}$, пока $0 \leqslant t < 2\pi$; кроме того, $e^{2\pi i} = 1$. При дальнейшем возрастании t движение точки e^{it} повторяется, ибо $e^{i(t+2\pi)} = e^{2\pi i}e^{it} = e^{it}$. Значит, при $2\pi < t < 4\pi$ дуга окружности $\mathbb{S}^1$, идущая в положительном направлении от 1 к e^{it}, имеет длину $t - 2\pi$, а потому $\arg e^{it} = t - 2\pi$, так что t — снова один из аргументов комплексного числа e^{it}, и т. д. (см. рис. 18). Сходные соображения применяются к отрицательным t (с увеличением такого t по абсолютной величине движение происходит в отрицательном направлении). Наконец, с одной стороны, ввиду (16) эйлеровские $\cos t$ и $\sin t$ суть декартовы координаты точки e^{it}, а с другой стороны — ввиду (2) такими координатами являются обычные $\cos\varphi$ и $\sin\varphi$, где φ — любой из аргументов точки e^{it}, т. е. её полярных углов. А t как раз и является одним из таких углов. Стало быть, в вещественной области эйлеровские тригонометрические функции и в самом деле совпадают с обычными.

После сказанного ясно, что (3) можно переписать в виде

$$r_1e^{i\varphi_1} \cdot r_2e^{i\varphi_2} = r_1r_2e^{i(\varphi_1+\varphi_2)}. \tag{18}$$

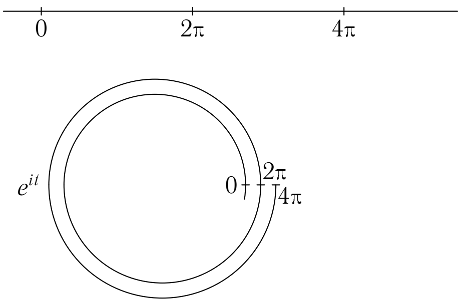


Рис. 18. При $t \to e^{it}$ прямая «наматывается» на окружность

Более того, этим фактически заново доказана формула (3), потому что выше она использовалась только один раз — когда мы говорили, что iz получается из z при повороте вокруг 0 в положительном направлении на прямой угол (и применяли это к $z = e^{it}$). Но это легко доказать непосредственно, а тогда отождествление t с одним из аргументов числа e^{it} и эйлеровских косинуса и синуса с обычными cos и sin получится независимо от (3). После отождествления формула Эйлера (13) сразу сводит (3) к (18).

Ясно также, что можно определить e^z следующим образом: если $z = x + iy$ с $x, y \in \mathbb{R}$, то $e^z = e^x(\cos y + i \sin y)$. Такое определение, не использующее степенных рядов, часто дают в тех случаях, когда хотят пораньше ввести экспоненту в комплексной области. При таком подходе доказательства основных свойств экспоненты получаются примерно столь же просто, как и выше[26)], но его недостатком является немотивированность исходного определения.

С другой стороны, различные тригонометрические формулы, имеющиеся в учебниках, получаются с помощью (14) и (15) и ещё «формул приведения», указывающих, что происходит с косинусом и синусом при изменении аргумента на кратное $\frac{\pi}{2}$. Формулы приведения следуют из второй и третьей формул (15) с использованием того, что

$$\cos\frac{\pi}{2} = 0, \quad \sin\frac{\pi}{2} = 1. \tag{19}$$

Отмечу без доказательства, что при эйлеровской трактовке тригонометрических функций (когда они определяются посредством рядов (12)) можно, не обращаясь к геометрии, доказать следующее: $\cos x$ обращается в нуль где-то на положительной вещественной полуоси $[0, \infty)$; значит, у него име-

[26)] Конечно, тогда доказательство (6) опирается на (3).

ется наименьший положительный нуль[27] α (используем непрерывность косинуса и то, что $\cos 0 = 1$); кроме того, оказывается, что $\sin\alpha = 1$ и $\sin x > 0$ на интервале $(0, 2\alpha)$. Если теперь принять за π число 2α, то (19) будет выполнено. При таком подходе к тригонометрическим функциям геометрия понадобится только тогда, когда мы захотим установить, какое отношение это новое π имеет к длине окружности или площади круга.

«Тригонометрия без геометрии» может показаться какой-то вычурной выдумкой, поскольку в элементарной математике тригонометрические функции не только определяются геометрически, но и используются исключительно или почти исключительно в геометрии. Однако в математическом анализе тригонометрические функции используются именно как функции, безотносительно к геометрии, и примерно столь же часто, как многочлены. С этой точки зрения геометрическое происхождение тригонометрических функций — чуть ли не историческая случайность, а тот факт, что они всё-таки используются в геометрии, объясняется формулой Эйлера (13) и тем, что t является одним из аргументов e^{it}.

* * *

Далее нам будет надо знать только то, что точки единичной окружности $\mathbb{S}^1$ суть точки (комплексные числа) вида $e^{it} = \cos t + i\sin t$, $t \in \mathbb{R}$, и что для них имеется операция умножения $(z, w) \mapsto zw$, по отношению к которой $\mathbb{S}^1$ является коммутативной группой[28]. В группе известным образом определяется ещё операция перехода к обратному элементу $z \mapsto z^{-1}$ (в случае $\mathbb{S}^1$ она нам уже известна). Обе эти операции ($(z, w) \mapsto zw$ и $z \mapsto z^{-1}$) в нашем случае непрерывны (в связи с этим говорят, что $\mathbb{S}^1$ является топологической группой).

Мы могли бы и не вводить комплексных чисел, а рассматривать в плоскости $\mathbb{R}^2$ единичную окружность $\mathbb{S}^1$ и отображение

$$\mathbb{R} \to \mathbb{S}^1, \quad t \mapsto (\cos t, \sin t).$$

Обозначив $(\cos t, \sin t)$ через $e(t)$, мы могли бы превратить $\mathbb{S}^1$ в коммутативную топологическую группу, положив $e(s)e(t) = e(s+t)$. Став на

[27] «Нулём» функции f называют те z, где $f(z) = 0$. (Это, пожалуй, своего рода математический жаргон, с которым читатель мог и не встречаться.)

[28] Группа — это множество G, в котором введена некоторая операция «умножения», сопоставляющая каждым двум элементам $z, w \in G$ некий третий элемент — их «произведение» zw; при этом должны выполняться определённые правила: ассоциативность $(z_1z_2)z_3 = z_1(z_2z_3)$; существование единицы, т. е. такого элемента 1, что $1z = z1$ при всех z; существование обратного элемента — для любого z имеется такой элемент z^{-1}, что $zz^{-1} = z^{-1}z = 1$. Что в нашем случае они выполняются, следует из уже известных нам свойств комплексных чисел. Коммутативность группы означает, что произведение не зависит от порядка сомножителей: $zw = wz$.

такой путь, мы должны были бы прежде всего убедиться в корректности определения, т. е. проверить, что при изменении t на кратные 2π результат не меняется. (При «комплексном» подходе такого вопроса не возникало — умножение определялось во всём $\mathbb{C}$ и это никак не было связано с представлением точек $\mathbb{S}^1$ в виде $e(t)$.) Далее, мы должны были бы сформулировать, каким правилам должно подчиняться это умножение в группе, и проверить, что в нашем случае они выполняются. (При «комплексном» подходе мы могли себе позволить игнорировать этот вопрос и даже не формулировать соответствующие правила, потому что они фактически содержатся в правилах, которым подчиняется умножение в $\mathbb{C}$ и которые можно считать привычными даже для тех читателей, которые слабо знакомы с $\mathbb{C}$, ибо эти правила аналогичны правилам, относящимся к умножению в $\mathbb{R}$.) Наконец, надо было бы проверить непрерывность отображений $(z, w) \mapsto zw$ и $z \mapsto z^{-1}$.

По существу, при таком подходе мы с самого начала отождествляем $\mathbb{S}^1$ с так называемой факторгруппой $\mathbb{R}/2\pi\mathbb{Z}$, где в качестве групповой операции в $\mathbb{R}$ используется сложение, $\mathbb{Z}$ — множество целых чисел, а $2\pi\mathbb{Z}$ — множество целых чисел, умноженных на 2π. Элементы факторгруппы суть множества, каждое из которых имеет вид $\{t + 2\pi n;\ n \in \mathbb{Z}\}$ с фиксированным $t \in \mathbb{R}$ (это так называемые смежные классы по подгруппе $2\pi\mathbb{Z}$), а групповая операция в ней (операцию в данном случае естественно записывать как сложение, ибо она получается из сложения в $\mathbb{R}$) определяется так:

$$\{t + 2\pi n;\ n \in \mathbb{Z}\} + \{s + 2\pi n;\ n \in \mathbb{Z}\} = \{t + s + 2\pi n;\ n \in \mathbb{Z}\}.$$

Проверка корректности определения сложения — это то же самое, что проверка корректности определения умножения $e(t)e(s) = e(t + s)$ в $\mathbb{S}^1$. Если можно так выразиться, факторгруппа дарит нам групповую операцию, а $\mathbb{S}^1$ — понятия типа непрерывности и предела.

Конечно, и не отождествляя $\mathbb{R}/2\pi\mathbb{Z}$ с $\mathbb{S}^1$, можно было бы проанализировать, как определить для нашей факторгруппы понятия типа непрерывности и предела, отправляясь от привычных понятий для $\mathbb{R}$. Но при таком пути надо стать на более абстрактную точку зрения в отношении подобных понятий. Ведь факторгруппа поначалу является просто некоторым множеством (множеством смежных классов); надо обсудить, что надо добавить к понятию множества, какую структуру надо в нём ввести, чтобы в нём можно было говорить о непрерывности и пределах. После того как это сделано, сам перенос соответствующих понятий из $\mathbb{R}$ в $\mathbb{R}/2\pi\mathbb{Z}$ выглядит довольно просто. Однако ради одного этого примера развивать целую систему абстрактных понятий не стоит, коль скоро обращение к $\mathbb{S}^1$ сразу доставляет нам всё необходимое и притом в достаточно привычном (или близком к привычному) виде. А раз мы всё-таки решили иметь дело с единичной окружностью, а не с факторгруппой, то писать $e(t)$, когда давным-давно имеется стандартное обозначение e^{it} — это было бы какой-то игрой в жмурки.

§ 3. Некоторые общие понятия и факты, связанные с непрерывностью

Часть математики, где обсуждается комплекс понятий, связанных с непрерывностью и пределами, называется «общей (или теоретико-множественной) топологией». Обсуждение приходится проводить в весьма общем и абстрактном виде, чтобы можно было охватить большое число конкретных примеров разнообразного характера. Нам же нужно лишь немногое, и то только применительно к подмножествам плоскости и, самое бо́льшее, иногда подмножествам $\mathbb{R}^n$. Более подготовленным читателям всё или почти всё из этого должно быть известно, кроме, может быть, теоремы о продолжении функции. Им достаточно бегло просмотреть этот параграф, останавливаясь на том немногом, что окажется для них новым.

Все рассматриваемые сейчас понятия определяются у нас в конечном счёте с помощью расстояния. (Когда мы раньше писали $|z - w|$, это было просто способом записи для $\rho(z, w)$.) Поэтому мы будем сразу говорить о формально несколько более общем случае арифметического пространства $\mathbb{R}^n$ и его подмножеств[29)]. Для $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ можно по аналогии с (1) положить

$$\rho(x, y) = \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}. \tag{20}$$

[29)] Читатель, знакомый (хотя бы немного) с метрическими пространствами, поймёт (если ещё не знает), что соответствующие определения дословно переносятся на этот случай.

С другой стороны, в основном нам нужен случай $\mathbb{R}^2$, и, при желании, читатель может им и ограничиться. Правда, есть важное исключение: нам порой приходится иметь дело с функциями (отображениями), зависящими от двух «переменных» x и y, принадлежащих некоторым подмножествам $A, B \subset \mathbb{R}^n$ (можно считать, что $n = 2$). Пары (x, y) образуют $A \times B$, так что в подобных случаях речь идёт об отображениях вида $f\colon A \times B \to C$. Поскольку $A \times B \subset \mathbb{R}^n \times \mathbb{R}^n = \mathbb{R}^{2n}$, то f является функцией (со значениями в C), определённой на некотором подмножестве $\mathbb{R}^{2n}$; соответственно понимается и её непрерывность. При $n = 2$ получается $\mathbb{R}^4$; надеюсь, что даже не очень подготовленный читатель этого не испугается. (Можно также напомнить, что непрерывность функции $g(x, y)$ двух вещественных переменных часто определяют, не обращаясь явно к $\mathbb{R}^2$, хотя с чуть более абстрактной точки зрения такое отдельное определение излишне — g является просто функцией, определённой на некотором подмножестве плоскости, и её непрерывность — это её непрерывность как функции от одной переменной — точки плоскости. Совершенно аналогично можно понимать и непрерывность отображений $f\colon A \times B \to C$, не привлекая явно $\mathbb{R}^{2n}$.)

Точно так же можно определить «длину» или «норму» $|\boldsymbol{u}|$ вектора $\boldsymbol{u} = (u_1, \ldots, u_n)$:

$$|\boldsymbol{u}| = \sqrt{u_1^2 + \ldots + u_n^2}.$$

Отметим, что

$$\rho(x, z) \leqslant \rho(x, y) + \rho(y, z)$$

(так называемое «неравенство треугольника» — в треугольнике с вершинами x, y, z длина стороны, соединяющей x с z, не превышает суммы длин других сторон). Это эквивалентно тому, что для векторов $\boldsymbol{u} = (u_1, \ldots, u_n)$, $\boldsymbol{v} = (v_1, \ldots, v_n)$

$$|\boldsymbol{u} + \boldsymbol{v}| \leqslant |\boldsymbol{u}| + |\boldsymbol{v}|. \tag{21}$$

Можно доказать, что в плоскости, содержащей точки x, y, z, справедлива обычная геометрия; значит, наше неравенство — просто одно из утверждений последней. Но такое доказательство требует больше места, чем обычно приводимое непосредственное доказательство неравенства (21) (мы ведь, собственно, даже не определили, что такое плоскость в $\mathbb{R}^n$, и не выяснили, действительно ли существует там плоскость, содержащая точки x, y, z). Это доказательство связано со следующим свойством скалярного произведения векторов, которое в данном случае есть $(\boldsymbol{u}, \boldsymbol{v}) = \sum\limits_{i=1}^{n} u_i v_i$:

$$|(\boldsymbol{u}, \boldsymbol{v})| \leqslant |\boldsymbol{u}||\boldsymbol{v}| \tag{22}$$

(неравенство Коши—Буняковского—Шварца, впрочем, известное ещё в XVIII в.). Доказательство формулы (22) таково: поскольку $|\boldsymbol{w}|^2 = (\boldsymbol{w}, \boldsymbol{w})$ и

$$(\boldsymbol{a} + \boldsymbol{b}, \boldsymbol{c} + \boldsymbol{d}) = (\boldsymbol{a}, \boldsymbol{c}) + (\boldsymbol{b}, \boldsymbol{c}) + (\boldsymbol{a}, \boldsymbol{d}) + (\boldsymbol{b}, \boldsymbol{d}),$$

то

$$0 \leqslant \left|\frac{\boldsymbol{u}}{|\boldsymbol{u}|} - \frac{\boldsymbol{v}}{|\boldsymbol{v}|}\right|^2 = \left(\frac{\boldsymbol{u}}{|\boldsymbol{u}|} - \frac{\boldsymbol{v}}{|\boldsymbol{v}|}, \frac{\boldsymbol{u}}{|\boldsymbol{u}|} - \frac{\boldsymbol{v}}{|\boldsymbol{v}|}\right) = \\ = \frac{(\boldsymbol{u}, \boldsymbol{u})}{|\boldsymbol{u}|^2} + \frac{(\boldsymbol{v}, \boldsymbol{v})}{|\boldsymbol{v}|^2} - 2\frac{(\boldsymbol{u}, \boldsymbol{v})}{|\boldsymbol{u}||\boldsymbol{v}|} = 2 - 2\frac{(\boldsymbol{u}, \boldsymbol{v})}{|\boldsymbol{u}||\boldsymbol{v}|},$$

$$\frac{(\boldsymbol{u}, \boldsymbol{v})}{|\boldsymbol{u}||\boldsymbol{v}|} \leqslant 1.$$

Здесь подразумевается, что $|\boldsymbol{u}| \neq 0$, $|\boldsymbol{v}| \neq 0$. Если же $|\boldsymbol{u}| = 0$ или $|\boldsymbol{v}| = 0$, то $\boldsymbol{u} = 0$ или $\boldsymbol{v} = 0$, и (22) очевидно. А теперь сразу получается и доказательство неравенства (21): оно эквивалентно тому, что

$$(\boldsymbol{u}+\boldsymbol{v},\, \boldsymbol{u}+\boldsymbol{v}) \leqslant (|\boldsymbol{u}|+|\boldsymbol{v}|)^2,$$

т. е. тому, что

$$|\boldsymbol{u}|^2+|\boldsymbol{v}|^2+2(\boldsymbol{u},\, \boldsymbol{v}) \leqslant |\boldsymbol{u}|^2+|\boldsymbol{v}|^2+2|\boldsymbol{u}||\boldsymbol{v}|.$$

В понятном смысле можно говорить о сходящейся последовательности точек $\mathbb{R}^n$ и о её пределе, равно как и о пределе функции, заданной на подмножестве $A \subset \mathbb{R}^n$, или отображения $f\colon A \to \mathbb{R}^m$, а также о непрерывности функции или подобного отображения. Во всех этих случаях определения аналогичны тем, которые были даны в случае комплексной плоскости $\mathbb{C}$ и её подмножеств.

Упражнение 7. Докажите непрерывность расстояния $\rho(x,\, y)$ как функции от двух переменных x, y.

Открытые и замкнутые множества

Множество $A \subset \mathbb{R}^n$ называется открытым, если вместе с каждой своей точкой оно содержит и все достаточно близкие к ней точки. Подробнее это означает: для любой точки $x \in A$ имеется такое число $\varepsilon > 0$, что все точки y, для которых $\rho(x,\, y) < \varepsilon$, тоже содержатся в A (см. рис. 19).

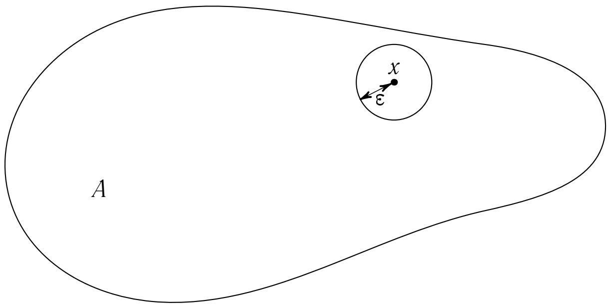


Рис. 19. Открытое множество

Упражнение 8. Докажите, что пересечение двух (а значит, и любого конечного числа) открытых множеств открыто и что объединение любой (даже бесконечной) системы открытых множеств открыто.

Упражнение 9. а) Докажите, что отображение $f\colon A \to B$ (где A, B — подмножества $\mathbb{R}^n$ и $\mathbb{R}^m$ соответственно) непрерывно в точке $x_0 \in A$ тогда и только тогда, когда для любого открытого множества $V \subset \mathbb{R}^m$, содержащего $f(x_0)$, найдётся такое открытое множество $U \subset \mathbb{R}^n$, содержащее x_0, что $f(U \cap A) \subset V$. Это отображение всюду непрерывно тогда и только тогда, когда для любого открытого множества $V \subset \mathbb{R}^m$ найдётся такое открытое множество $U \subset \mathbb{R}^n$, что прообраз [30] $f^{-1}(V)$ содержится в U.

б) Сформулируйте в терминах открытых множеств определение предела последовательности или предела отображения.

Упражнение 9 можно рассматривать как формулировку понятий непрерывности и предела, которая внешне выглядит более геометрической, чем обычная формулировка из курса анализа, где пишутся неравенства. Однако и обычной формулировке ничего не стоит придать геометрический характер, например: «отображение $f\colon A \to B$ называется непрерывным в точке $x_0 \in A$, если для любого открытого шара [31] с центром в $f(x_0)$ найдётся такой открытый шар с центром в x_0, пересечение которого с A под действием f отображается в первый шар». Только для непрерывности отображения во всех точках A формулировка с открытыми множествами действительно выглядит новой (в ней ведь не говорится о поведении f возле отдельных точек).

Если мы сосредоточили внимание на некотором множестве $A \subset \mathbb{R}^n$, которое на какое-то время образует для нас как бы «весь наш мир», то полезно ввести относительные понятия, которые связаны именно с A и при обращении к которым мы как бы забываем о существовании объемлющего пространства $\mathbb{R}^n$. Некоторые понятия (например, сходимость последовательности точек $x_n \in A$ к точке $a \in A$; непрерывность заданной на A функции f в какой-нибудь точке множества A или во всех его точках) по существу с самого начала таковы — $\mathbb{R}^n$ фигурирует в соответствующих определениях лишь посредством понятия расстояния, причём используется только расстояние между точками A. Относительные варианты других понятий связаны с небольшими модификациями последних. Открытым шаром радиуса $r > 0$ в A с центром в точке $a \in A$ естественно называть $\{y;\ y \in A,\ \rho(y, a) < r\}$. Он является пересечением с A соответствующего

---

[30] То есть множество тех $x \in A$, образы которых при отображении f попадают в V:

$$f^{-1}(V) = \{x;\ x \in A,\ f(x) \in V\}.$$

(Запись вида

$$\{x;\ x \text{ удовлетворяет таким-то условиям}\}$$

означает множество всех тех x, для которых выполняются указанные условия.) Прообраз $f^{-1}(y)$ элемента y — это то же самое, что и прообраз $f^{-1}(\{y\})$ «одноточечного» множества $\{y\}$ (множества, состоящего из одного-единственного элемента y), т. е. множество тех $x \in A$, для которых $f(x) = y$.

Наряду со сказанным, под прообразом элемента y часто понимают любой из элементов множества $f^{-1}(y)$ (о последнем в таком случае говорят как о полном прообразе).

[31] Открытый шар пространства $\mathbb{R}^m$ с центром в точке $y_0 \in \mathbb{R}^m$ — это множество $\{y;\ y \in \mathbb{R}^m,\ \rho(y, y_0) < r\}$. (Проверьте, используя неравенство треугольника, что это действительно открытое множество.) Множество $\{y;\ y \in \mathbb{R}^m,\ \rho(y, y_0) \leqslant r\}$ естественно называть замкнутым шаром.

открытого шара пространства $\mathbb{R}^n$. Множество $B \subset A$ называется относительно открытым (подразумевается: открытым относительно A) или открытым в A, если вместе с каждой своей точкой x оно содержит все точки некоторого открытого шара в A с центром в x[32].

Упражнение 10. Докажите, что подмножество $B \subset A$ тогда и только тогда открыто в A, когда оно является пересечением с A некоторого открытого подмножества пространства $\mathbb{R}^n$.

Теперь можно сократить формулировку упражнения 9, говоря в нём об относительно открытых множествах. А тогда получается такая переформулировка определения гомеоморфизма $f\colon A \to B$: это такое взаимно однозначное отображение A на B, при котором образы и прообразы открытых в A или B подмножеств суть открытые в B или A подмножества. Иными словами, при нашем взаимно однозначном соответствии между множествами A и B получается и взаимно однозначное соответствие между системой всех относительно открытых подмножеств A и системой всех относительно открытых подмножеств B. Если «заменить» A на B таким образом, что каждый элемент $x \in A$ заменится на $f(x) \in B$, то это никак не отразится на системе относительно открытых подмножеств — относительно открытое подмножество останется относительно открытым, а новых относительно открытых подмножеств при такой замене не получится.

Коль скоро гомеоморфизмы так просто определяются с помощью относительно открытых множеств, то не удивительно, что топологические свойства множества A в конечном счёте определяются системой его относительно открытых подмножеств. Можно изменить расстояния между точками, скажем, увеличить их в два раза — открытые множества при этом останутся теми же. Менее тривиальное изменение получается, если условиться понимать под расстоянием между точками $x = (x_1, \ldots, x_n)$, $y = (y_1, \ldots, y_n) \in \mathbb{R}^n$ некоторую функцию от x, y, отличающуюся от (20) и не имеющую непосредственного геометрического аналога, но всё-таки как-то характеризующую взаимное удаление этих точек. Может случиться, что при замене в определении открытых множеств прежнего расстояния на эту новую функцию система открытых подмножеств $\mathbb{R}^n$ не изменится, а тогда не изменятся и все топологические свойства любых подмножеств $\mathbb{R}^n$. Например, можно взять

$$\rho_1(x, y) = \max(|x_1 - y_1|, \ldots, |x_n - y_n|)$$

или

$$\rho_2(x, y) = |x_1 - y_1| + \ldots + |x_n - y_n|.$$

[32] Читатель, знакомый с понятием метрического пространства, заметит, что наши относительные понятия суть в точности соответствующие первоначальные понятия для метрического пространства (A, ρ), т. е. для A, рассматриваемого вместе с тем расстоянием между его точками, которое является их расстоянием в $\mathbb{R}^n$.

Для таких расстояний тоже выполняется неравенство треугольника, и оно получается проще, чем для (20) (проверьте).

Упражнение 11. Покажите, что при использовании ρ_1 или ρ_2 вместо (20) в определении открытого подмножества $\mathbb{R}^n$ система открытых подмножеств не меняется.

Хотя топологические свойства множества A в конечном счёте определяются системой его относительно открытых подмножеств, отнюдь не очевидно, что стоит принять точку зрения, при которой эта система ставится, так сказать, во главе угла. Учащемуся это так же не очевидно, как это было не очевидно выдающимся учёным в начале XX в., которые пробовали различные другие варианты построения топологии. Во всех вариантах расстояния могли использоваться только при определении какого-то исходного класса объектов, да и то в иных случаях эти объекты могли вводиться и как-нибудь иначе, в дальнейшем же всё делалось на основании одного этого класса. Практика первой четверти века показала, что вариант с открытыми множествами всего лучше. С тех пор общепринятым стало понятие «топологического пространства» — множества A вместе с некоей системой его подмножеств, именуемых открытыми, причём данная система должна удовлетворять нескольким простым условиям[33]. Мы не будем вникать в эти абстрактные дебри — наши топологические пространства будут подмножествами $\mathbb{R}^n$, для которых нужные нам понятия (отчасти, вероятно, уже известные читателю) вводятся без обращения к подобного рода общей концепции[34]. Но (относительно) открытые множества будут играть у нас большую роль.

С геометрической точки зрения, когда мы говорим о малости расстояний до фиксированной точки a, мы фактически имеем дело с шарами малого радиуса с центром в a. Но шары — это не топологическое понятие. Можно было бы подумать, что с топологической точки зрения надо говорить о множествах, гомеоморфных шарам. Однако в вопросах, связанных с непрерывностью, пределами и т. п. оказалось целесообразным заменять малые шары гораздо более широким классом множеств — так называемыми окрестностями точки a. Под открытой окрестностью точки A понимают

---

[33] Главные из них указаны в упражнении 8; сверх того требуется, чтобы пустое множество и всё множество A считались открытыми. Для начала этим ограничиваются, но позднее появляются различные дополнительные условия, обеспечивающие наличие у топологического пространства различных «хороших» свойств, существенных в приложениях.

[34] Бывает, что несколько вариантов того или иного определения или утверждения, эквивалентных применительно к подмножествам $\mathbb{R}^n$, в общем случае не эквивалентны. При разработке общей теории пришлось думать о рациональном выборе вариантов (что не всегда было просто). При изучении этой теории приходится загрузить в память некоторое количество окончательно отобранных вариантов и понимать, почему решено было остановиться именно на них (а также когда они эквивалентны другим). Имея дело только с подмножествами $\mathbb{R}^n$, мы свободны от таких забот.

любое открытое множество, содержащее эту точку, а любое множество, содержащее открытую окрестность точки a, называют просто окрестностью этой точки. (Впрочем, нередко под окрестностями понимают только открытые окрестности.) В терминах расстояния (которое для нас всё-таки было исходным понятием) окрестность a — это множество, содержащее некоторый шар с центром в a. Далее, «оставаясь в пределах» какого-нибудь множества $A \subset \mathbb{R}^n$, говорят об относительных окрестностях его точки a (её окрестностях в A). Относительная окрестность точки a — это множество, содержащее некоторое открытое в A множество, которое в свою очередь содержит a. Она же содержит некоторый шар в A с центром a и является пересечением с A некоторой окрестности точки a в объемлющем пространстве $\mathbb{R}^n$.

Всё это пока что — не больше, чем названия. Но опыт показывает, что в терминах окрестностей получаются более короткие перефразировки различных определений и рассуждений. Например: отображение $f\colon A \to B$ непрерывно в точке a, если прообраз любой окрестности V точки $f(a)$ в B является окрестностью a в A, т. е. если в A у a имеется такая окрестность U, что $f(U) \subset V$.

Упражнение 12. Перефразируйте в терминах окрестностей определения предела последовательности точек или отображения.

Упражнение 13. Докажите, что пересечение двух окрестностей данной точки — окрестность. Отсюда следует, что пересечение конечной системы окрестностей — окрестность. Верно ли это для бесконечной системы?

Точка a называется предельной для множества A (или предельной точкой этого множества), если она является пределом некоторой последовательности точек этого множества, отличных от a. Эквивалентное определение: a — предельная точка множества A, если сколь угодно близко к ней (т. е. в любой её окрестности) имеется точка множества A, отличная от a. Чуть более широкое понятие — точка прикосновения множества A — получается, если отбросить оговорку «отличная от». Если точка $a \in A$, то, конечно, в любой её окрестности имеется точка из A — сама a. Если же точка прикосновения $a \notin A$, то близкие к ней точки A неизбежно отличны от a, и тогда a — предельная точка множества A. Итак, точка прикосновения множества A — это или точка из A, или точка, предельная для A, или и то, и другое. Можно ещё сказать, что это предел некоторой последовательности x_n точек из A (здесь не исключено, что некоторые или даже все x_n равны a). Точка множества A, не являющаяся предельной для A, называется изолированной точкой этого множества.

Замыкание $\operatorname{clos} A$ (пишут также $\overline{A}$) множества A — это множество всех его точек прикосновения, т. е. это само A вместе со всеми его предельными точками. Множество называется замкнутым, если оно совпадает со своим замыканием. Имея дело с подмножествами одного и того же множества A, вводят соответствующие относительные понятия. Точка прикосновения такого подмножества B в A — это просто точка прикосновения, лежащая в A (в её определении можно говорить об относительных окрестностях). Замыкание B в A — это множество всех точек прикосновения B в A, т. е. $\operatorname{clos} B \cap A$. B замкнуто в A (или замкнуто относительно A), если B совпадает со своим замыканием в A, т. е. $B = \operatorname{clos} B \cap A$.

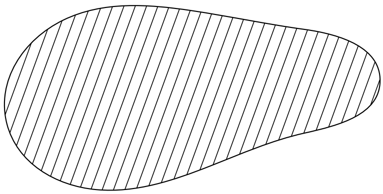


Рис. 20. Замкнутое множество

Упражнение 14. а) Докажите, что замкнутые подмножества $\mathbb{R}^n$ суть в точности дополнения (в $\mathbb{R}^n$) к открытым (т. е. если A открыто, то $\mathbb{R}^n \setminus A$ замкнуто, и обратно).

б) Докажите, что пересечение любой (даже бесконечной) системы замкнутых множеств замкнуто и что объединение двух или конечного числа замкнутых множеств замкнуто. Обязательно ли будет замкнутым объединение бесконечного числа замкнутых множеств? (См. рис. 21.)

в) Докажите, что $\operatorname{clos} A$ — наименьшее замкнутое множество, содержащее A.

г) Сформулируйте и проверьте относительные аналоги этих утверждений.

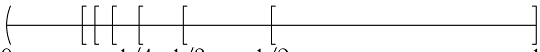


Рис. 21. $(0,\, 1] = \bigcup \left[\dfrac{1}{n},\, 1\right]$

Граница $\operatorname{Fr} A$ множества A определяется так:

$$\operatorname{Fr} A = \{x;\ \text{в любой окрестности } U \text{ точки } x \text{ имеются как точки, принадлежащие } A, \text{ так и точки, не принадлежащие } A\}.$$

(Fr происходит от французского frontière.) Из определения видно, что

$$\operatorname{Fr} A = \operatorname{clos} A \cap \operatorname{clos}(\mathbb{R}^2 \setminus A).$$

В частности, $\operatorname{Fr} A$ — замкнутое множество.

Компактность

В курсах анализа доказывается, что из каждой ограниченной последовательности вещественных чисел можно выделить сходящуюся подпоследовательность. Отсюда легко вывести аналогичное утверждение об ограниченной последовательности точек $x^{(j)} \in \mathbb{R}^n$ (её ограниченность означает, что все $|x^{(j)}|$ не превосходят некоторого вещественного числа C, не зависящего от j). Действительно, координаты точек $x^{(j)}$ тем более не превосходят C по модулю. Поэтому из $\{x^{(j)}\}$ можно выбрать такую подпоследовательность точек, что их первые координаты образуют сходящуюся последовательность. В этой подпоследовательности содержится такая (ещё более «разреженная») подпоследовательность точек, что вторые координаты этих точек тоже образуют сходящуюся последовательность. Продолжая этот процесс, на k-м шаге получим подпоследовательность точек, для которых сходятся последовательности их первых, вторых, ..., k-х координат. При $k = n$ все координаты точек соответствующей подпоследовательности образуют сходящиеся последовательности, что и означает сходимость n-й подпоследовательности.

Множество $A \subset \mathbb{R}^n$ называется компактным[35], если в любой последовательности его точек содержится сходящаяся в A подпоследовательность (т. е. сходящаяся подпоследовательность, предел которой принадлежит A).

Теорема. *Компактные подмножества пространства $\mathbb{R}^n$ суть в точности замкнутые ограниченные[36] подмножества этого пространства[37].*

Доказательство. Пусть A компактно. Тогда предел $\lim x_k$ любой сходящейся последовательности точек из A принадлежит A, потому что из неё можно извлечь сходящуюся в A подпоследовательность, а та может сходиться только к $\lim x_k$. Значит, A замкнуто. Далее, допустим, что A не ограничено. Тогда для любого натурального числа m в A найдётся точка x_m с $|x_m| \geqslant m$. В последовательности $\{x_m\}$ не содержится никакой сходящейся подпоследовательности, вопреки компактности A.

Обратно, пусть A — ограниченное замкнутое подмножество $\mathbb{R}^n$ и $\{x_j\}$ — последовательность точек из A. Она тоже является ограниченной; значит, у неё имеется сходящаяся подпоследовательность $\{x_{j_k}\}$. Предел последней принадлежит множеству A ввиду замкнутости этого множества.

Для непрерывных функций на компактном множестве K справедливы теоремы, совершенно аналогичные известным теоремам о функциях, непрерывных на замкнутом отрезке.

Теорема. а) *Непрерывная функция $f\colon K \to \mathbb{R}$ является ограниченной (т. е. имеется такое число C, что $|f(x)| < C$ при всех $x \in K$).*

---

[35] Данное определение годится и для более общих метрических пространств. В ещё более общем случае топологических пространств компактность определяется иначе.

[36] Множество $A \subset \mathbb{R}^n$ называется ограниченным, если имеется такое число C, что $|x| < C$ для всех $x \in A$.

[37] Для общих метрических пространств это не так.

б) Она принимает где-то на K максимальное и минимальное значения, т. е. имеются такие $x_{\max}$, $x_{\min} \in K$, что

$$f(x_{\min}) \leqslant f(x) \leqslant f(x_{\max}) \quad \textit{при всех} \;\; x \in K.$$

Доказательство. Допустим, что f не ограничена. Тогда для любого натурального числа j существует точка x_j, где $|f(x_j)| \geqslant j$. Пусть $\{x_{j_k}\}$ — сходящаяся подпоследовательность последовательности $\{x_j\}$ и $a = \lim\limits_k x_{j_k}$. Когда точка $x \in K$ лежит в некоторой окрестности U точки a, то $|f(x) - f(a)| < 1$. С другой стороны, $x_{j_k} \in U$ при достаточно больших k, — скажем, при $k > k_0$. Поэтому

$$j_k \leqslant |f(x_{j_k})| < |f(a)| + 1 \quad \text{при} \;\; k > k_0.$$

Но это невозможно — ведь $j_k \to \infty$ при $k \to \infty$. Полученное противоречие доказывает ограниченность f.

Ввиду этой ограниченности, верхняя грань $\sup\limits_{x\in K} f(x) < \infty$. Для любого натурального j имеется точка $x_j \in K$, где $f(x_j) > \sup\limits_{x\in K} f(x) - \dfrac{1}{j}$. Пусть $\{x_{j_k}\}$ — сходящаяся подпоследовательность последовательности $\{x_j\}$ и $a = \lim\limits_k x_{j_k}$. Тогда $f(x_{j_k}) \to f(a)$ при $k \to \infty$, поэтому $f(a) \geqslant \sup\limits_{x\in K} f(x)$, а так как $>$ здесь невозможно по смыслу понятия точной верхней грани, то $f(a) = \sup\limits_{x\in K} f(x)$. Значит, a годится в качестве $x_{\max}$. Для минимума рассуждения аналогичны. (Или можно рассмотреть максимум функции $-f(x)$.)

«Склеивание» непрерывных отображений из непрерывных «кусков»

Пусть множество $A = A_1 \cup A_2$ и на его подмножествах A_1, A_2 заданы отображения в некоторое множество B:

$$f_1\colon A_1 \to B, \qquad f_2\colon A_2 \to B$$

(см. рис. 22). В каком случае существует отображение $f\colon A \to B$, при котором образы точек A_1 — те же, что при f_1, а образы точек A_2 — те же, что при f_2, т. е. f_1 является ограничением[38] $f|A_1$ отображения f на A_1, а $f_2 = f|A_2$? Очевидно, f существует тогда и только тогда, когда f_1 и f_2 одинаково отображают точки $A_1 \cap A_2$, т. е. $f_1|A_1 \cap A_2 = f_2|A_1 \cap A_2$. (В таком случае безразлично, скажем ли мы, что для $x \in A_1 \cap A_2$ за образ $f(x)$ принимается $f_1(x)$, или $f_2(x)$. Для $x \in A_1 \setminus A_2$, соответственно для $x \in A_2 \setminus A_1$,

[38] Если имеется отображение $g\colon A \to B$, то ограничение (или сужение) $g|C$ этого отображения на подмножество $C \subset A$ — это то же самое отображение g, рассматриваемое только в точках C (т. е., говоря более формально, $g|C$ есть отображение $C \to B$, для которого $(g|C)(x) = f(x)$ при всех $x \in C$).

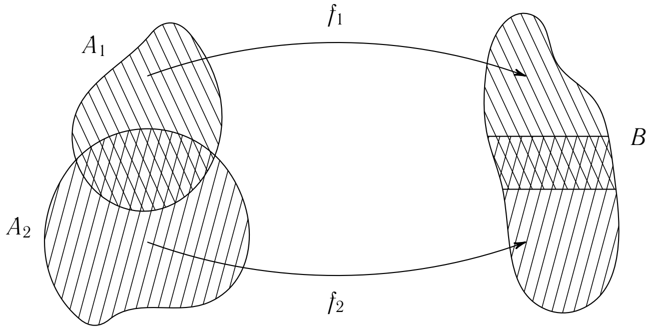


Рис. 22

вопросов нет: там $f(x) = f_1(x)$, соответственно $f(x) = f_2(x)$.) При выполнении последнего условия естественно говорить, что f_1 и f_2 «согласованы» (друг с другом) и что они «склеиваются» в отображение f.

Пусть теперь $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$, $A = A_1 \cup A_2$ и имеются непрерывные отображения $f_1 \colon A_1 \to B$, $f_2 \colon A_2 \to B$, которые «согласованы» в только что указанном смысле и, значит, «склеиваются» в отображение $f \colon A \to B$. Будет ли f непрерывно?

Не обязательно. Пусть, например,

$$A = [-1,\, 1], \quad A_1 = [-1,\, 0], \quad A_2 = (0,\, 1],$$
$$B = \mathbb{R}, \quad f|A_1 = 0 \ \text{ и } \ f|A_2 = 1$$

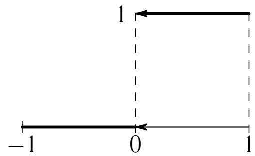


Рис. 23

(см. рис. 23.) В этом случае $f|A_1$ и $f|A_2$ непрерывны и они тривиальным образом согласованы друг с другом (ибо $A_1 \cap A_2 = \varnothing$, т. е. нет таких точек, где мы требовали бы совпадения f_1 с f_2), а f — разрывная функция.

Всё же имеются два простых и полезных случая, когда непрерывность f гарантирована.

1. *Если A_1 и A_2 замкнуты в A, то отображение f непрерывно.* Иными словами,

> если непрерывны ограничения $f|A_i$ отображения $f \colon A \to B$ на замкнутые в A подмножества A_1, A_2, объединение которых есть всё A, то f непрерывно. (23)

Действительно, непрерывность в каждой точке $x_0 \in A_1 \setminus A_2$ явствует из того, что у этой точки имеется окрестность в A, не пересекающаяся с замкнутым в A множеством A_2. Когда для точек x из этой окрестности мы сравниваем $f(x)$ с $f(x_0)$, то это то же самое, что сравнивать $f_1(x)$ и $f_1(x_0)$, а о последних точках нам дано, что при достаточной близости x к x_0 они будут близки. Аналогично получается непрерывность f в точках $A_2 \setminus A_1$. Наконец, если $x_0 \in A_1 \cap A_2$ и V — окрестность точки $f(x_0)$, то у x_0 имеются такие две окрестности W_1, W_2 в A, что когда $x \in W_1 \cap A_1$ или $x \in W_2 \cap A_2$, то $f(x)$ (совпадая, соответственно, с $f_1(x)$ или $f_2(x)$) будет лежать в V. Тогда $W = W_1 \cap W_2$ есть окрестность x_0 в A, для которой $f(W) \subset V$.

2. *Если A_1 и A_2 открыты в A, то отображение f непрерывно.* Это сразу сводится к аналогу (23), в котором A_i предполагаются открытыми в A. Доказательство непрерывности f в этом случае даже чуть проще и предоставляется читателю.

Теорема о продолжении функции

Теорема о продолжении функции. *Пусть на компактном подмножестве $K \subset \mathbb{R}^n$ задана непрерывная функция $f\colon K \to \mathbb{R}$. Тогда её можно продолжить до непрерывной функции на всём $\mathbb{R}^n$, т. е. существует такая непрерывная функция $F\colon \mathbb{R}^n \to \mathbb{R}$, ограничение которой на K есть f.*

Данная теорема справедлива и в более общем случае, когда K является замкнутым подмножеством $\mathbb{R}^n$ (не обязательно компактным)[39].

Одно из упрощений в нашем случае состоит в том, что ввиду компактности K и непрерывности f существуют m и M со свойством $m \leqslant f < M$ всюду. Добавив к f подходящую константу, можем считать $m > 0$. (Если мы сумеем продолжить $f(x) + \mathrm{const}$, то, вычтя из продолжения ту же самую константу, получим продолжение для f.) Положим

$$F(x) = \begin{cases} f(x), & \text{когда } x \in K, \\ \dfrac{\min\limits_{z \in K}(f(z)\rho(x, z))}{\rho(x, K)}, & \text{когда } x \in \mathbb{R}^n \setminus K. \end{cases}$$

Здесь $\rho(x, K)$ — расстояние от точки x до K; оно определяется как $\min\limits_{w \in K} \rho(x, w)$. Этот минимум действительно достигается в некоторой точке $z_x \in K$, поскольку $z \mapsto \rho(x, z)$ — непрерывная функция на компактном множестве K.

[39] При $n = 2$ такая теорема была доказана А. Лебегом, затем Л. Брауэр отметил, что она справедлива при любом n, а А. Титце — что она справедлива, когда K является замкнутым подмножеством любого метрического пространства, на которое и продолжается f. Во всех этих случаях доказательство аналогично тому, которое приводится у нас, только у нас компактность K позволила кое-что упростить. В ещё более общем случае, когда речь идёт о замкнутом подмножестве топологического пространства, продолжение не всегда возможно, но П. С. Урысон доказал, что оно возможно, когда пространство обладает некоторым «хорошим» свойством (так называемой «нормальности»), выполненным в большинстве приложений. В этом случае доказательство совсем другое.

Докажем, что определённая таким способом на всём $\mathbb{R}^n$ функция F, совпадающая с f на K, непрерывна.

Расстояние от x до K является непрерывной функцией от x; можно даже утверждать, что

$$|\rho(x, K) - \rho(y, K)| \leqslant \rho(x, y). \tag{24}$$

В самом деле,

$$\rho(x, K) - \rho(y, K) = \rho(x, z_x) - \rho(y, z_y) \leqslant \rho(x, z_y) - \rho(y, z_y) \leqslant \rho(x, y).$$

Первое неравенство здесь следует из $\rho(x, z_x) \leqslant \rho(x, z_y)$, а второе — из неравенства треугольника. Итак,

$$\rho(x, K) - \rho(y, K) \leqslant \rho(x, y), \tag{25}$$

а меняя здесь местами x и y, получаем, что

$$\rho(y, K) - \rho(x, K) \leqslant \rho(y, x) = \rho(x, y). \tag{26}$$

Неравенства (25) и (26) влекут (24).

То, что $F|K = f$, ясно из определения. Докажем непрерывность F — сперва в точках $\mathbb{R}^n \setminus K$, затем в точках K.

Непрерывность в точках $\mathbb{R}^n \setminus K$ эквивалентна непрерывности числителя в формуле для F, поскольку знаменатель непрерывен и не обращается в этих точках в 0. Итак, надо доказать непрерывность функции $G(x) = \min\limits_{z \in K}(f(z)\rho(x, z))$. При любом $z \in K$

$$|f(z)\rho(x, z) - f(z)\rho(y, z)| \leqslant |f(z)||\rho(y, z) - \rho(x, z)| \leqslant M\rho(x, y)$$

(почему $|\rho(y, z) - \rho(x, z)| \leqslant \rho(x, y)$?). Раз две функции от z при всех z различаются не более чем на некоторое число, их минимумы тоже различаются не более чем на это число (почему?). Вот и получается, что $|G(y) - G(x)| \leqslant M\rho(x, y)$, а это гарантирует непрерывность $G(x)$.

Пусть теперь $x \in K$ (так что $F(x) = f(x)$). Зафиксировав $\varepsilon > 0$, мы хотим доказать существование такого $\delta > 0$, что если $\rho(x, y) < \delta$, то $|f(x) - F(y)| < \varepsilon$. Поскольку f — непрерывная функция на K, то существует такое $\delta_1 > 0$, что когда $y \in K$ и $\rho(x, y) < \delta_1$, то $|f(x) - f(y)| < \varepsilon$ (а ведь для таких y, по определению, $F(y) = f(y)$). Сейчас мы докажем существование такого $\delta_2 > 0$, что когда $y \notin K$ и $\rho(x, y) < \delta_2$, то $|f(x) - F(y)| < \varepsilon$. Тогда в качестве δ можно взять $\min(\delta_1, \delta_2)$.

Итак, теперь $y \notin K$. Пусть по-прежнему z_y — ближайшая к y точка K. Тогда $\rho(y, z_y) = \rho(y, A) \leqslant \rho(x, y)$,

$$\min_{z\in K}(f(z)\rho(y, z)) \leqslant f(z_y)\rho(y, z_y) = f(z_y)\rho(y, K),$$

$$F(y) \leqslant f(z_y). \tag{27}$$

Но $\rho(x, z_y) \leqslant \rho(x, y) + \rho(y, z_y) \leqslant 2\rho(x, y)$. Пусть δ_3 — столь малое положительное число, что когда $z \in K$ и $\rho(x, z) < 2\delta_3$, то $|f(x) - f(z)| < \varepsilon$. Тогда при $y \notin K$, $\rho(x, y) < \delta_3$, будет $F(y) < f(x) + \varepsilon$.

Остаётся убедиться в верности неравенства $F(y) > f(x) - \varepsilon$ при достаточной близости y к x. Сперва мы проверим, что когда берётся минимум $\min\limits_{z\in K}(f(z)\rho(y, z))$, а y близко к x, то можно ограничиться только достаточно близкими к x точками $z \in K$. При любом $w \in K$

$$F(y) = \frac{\min\limits_{z\in K}(f(z)\rho(y, z))}{\rho(y, K)} \geqslant m\frac{\rho(y, w)}{\rho(y, K)},$$

и если $\rho(y, w) \geqslant \dfrac{M}{m}\rho(y, K)$, то $F(y) > M$. Но ведь (см. (27)) $F(y) \leqslant f(z_y) < M$. Значит, $\min$ в определении $F(y)$ не может достигаться на таких z, для которых $\rho(y, z) \geqslant \dfrac{M}{m}\rho(y, K)$. Он достигается на тех $z \in K$, для которых

$$\rho(y, z) < \frac{M}{m}\rho(y, K) \leqslant \frac{M}{m}\rho(x, y).$$

Для них тем более $\rho(x, z) \leqslant \rho(x, y) + \rho(y, z) < \dfrac{M+m}{m}\rho(x, y)$. Итак,

$$F(y) = \min_{z\in K,\ \rho(x,z)<\frac{M+m}{m}\rho(x,y)}\left(\frac{f(z)\rho(y, z)}{\rho(y, A)}\right) \geqslant \min_{z\in K,\ \rho(x,z)<\frac{M+m}{m}\rho(x,y)} f(z).$$

Пусть $\delta_4 > 0$ столь мало, что из $z \in K, \rho(x, z) < \dfrac{M+m}{m}\delta_4$ следует $|f(x) - f(z)| < \varepsilon/2$. Тогда при $\rho(x, y) < \delta_4$ в последнем минимуме фигурируют $z \in K$, для которых $\rho(x, z) < \dfrac{M+m}{m}\delta_4$, и потому этот минимум $\geqslant f(x) - \varepsilon/2 > f(x) - \varepsilon$. Окончательно получается, что в качестве искомого δ_2 годится $\min(\delta_3, \delta_4)$.

Дополнение к теореме о продолжении. *Пусть на компактном подмножестве $K \subset \mathbb{R}^n$ задана непрерывная функция $f\colon K \to \mathbb{R}$, и пусть U — открытое множество, содержащее K. Тогда существует непрерывное продолжение F функции f на всё $\mathbb{R}^n$, равное нулю всюду вне U.*

Можно считать, что множество U — ограниченное. Действительно, K является ограниченным множеством и потому целиком содержится в некотором открытом шаре B. Если дополнение будет доказано для ограниченных U, то можно применить его, взяв вместо U открытое множество $U \cap B$, которое тоже содержит K. Соответствующее непрерывное продолжение F будет равно нулю всюду вне $U \cap B$, а тем более вне U.

Когда U — ограниченное множество, ограниченным является и его замыкание $\operatorname{clos} U$, которое, стало быть, целиком содержится в некотором открытом шаре B. Множество $\operatorname{clos} B \setminus U$ — замкнутое. Положим $K_1 = K \cup (\operatorname{clos} B \setminus U)$ и

$$f_1(x) = \begin{cases} f(x) & \text{при } x \in K, \\ 0 & \text{при } x \in \operatorname{clos} B \setminus U. \end{cases}$$

Ясно, что K_1 — замкнутое ограниченное множество, так что оно компактно. Поскольку его подмножества K и $B \setminus U$ суть непересекающиеся замкнутые множества, то функция f_1 корректно определена и непрерывна.

Применим теорему о продолжении к K_1 вместо K и к f_1 вместо f. Получим, что существует непрерывная функция $F_1 \colon \mathbb{R}^n \to \mathbb{R}$, совпадающая с f_1 на K_1. Теорема ничего не говорит о её значениях вне B, поэтому изменим её вне B, положив

$$F(x) = \begin{cases} F_1(x) & \text{при } x \in B, \\ 0 & \text{при } x \notin B. \end{cases}$$

Из определения видно, что $F = 0$ всюду вне U: вне B мы только что приняли её равной нулю, а на $B \setminus U$ она совпадает с функцией F_1, которая в свою очередь совпадает там с f_1 и потому равна нулю. Поскольку $K \subset B$, то на $F|K = F_1|K = f$. Остаётся убедиться в непрерывности F. Заметим, что на самом деле $F = F_1$ не только на B, но и на $\operatorname{clos} B$, ибо в точках $\operatorname{clos} B \setminus B$, к которым первая строчка в определении F не относится, верно равенство $F = 0 = F_1$. Первое равенство ($F = 0$) видно из второй строчки определения, F_1 же в этих точках совпадает с функцией f_1, которая в этих точках равна нулю; то и другое явствует из того, что $\operatorname{clos} B \setminus B \subset \operatorname{clos} B \setminus U$. Раз $F|\operatorname{clos} B = F_1|\operatorname{clos} B$, то ограничение F на замкнутое множество $A_1 = \operatorname{clos} B$ непрерывно. Ограничение F на замкнутое множество $A_2 = \mathbb{R}^n \setminus U$ тоже непрерывно (будучи нулём). Таким образом, F «склеивается» из согласованных друг с другом функций $A_i \to \mathbb{R}$, определённых и непрерывных на замкнутых подмножествах $\mathbb{R}^n$, что, как мы знаем, обеспечивает её непрерывность.

Связность

Первоначальное «интуитивное» и наглядное представление о связности таково: фигура несвязна, если она, подобно букве Ы, состоит из двух отдельных, не связанных между собой частей; в противном случае фигура называется связной. Несомненно, буква И — связная фигура. Но является ли связной фигура F (если её можно назвать фигурой), состоящая из графика функции $y = \sin\frac{1}{x}$, $x > 0$, и вертикального отрезка $x = 0$, $-1 \leqslant y \leqslant 1$ (см. рис. 24)? Эти две части — график и вертикальный отрезок — не так

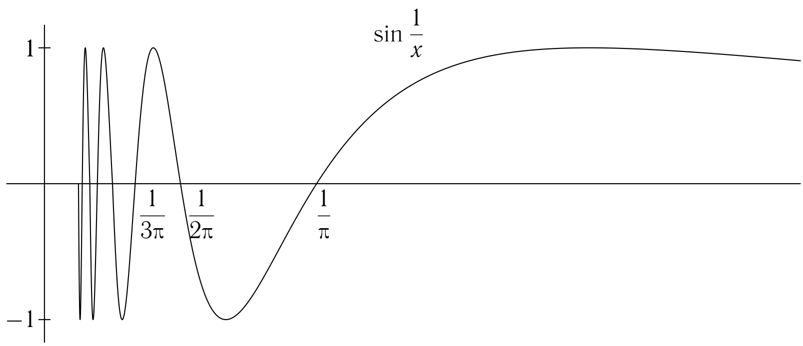


Рис. 24

сильно отделены друг от друга, как части Ь и I фигуры Ы: вертикальный отрезок лежит в замыкании графика. И всё же это вроде бы отдельные части, во всяком случае, двигаясь по фигуре F, нельзя перейти из графика в вертикальный отрезок. Можно сделать вывод, что фигура F менее связна, чем буква И, но более связна, чем буква Ы; получается, что связность (в отличие от свежести, — впрочем, это мнение Воланда) бывает разных сортов. Идея одной «разновидности» связности — та, с которой мы начали («две отдельные части»), идея другой — от одной точки фигуры можно перейти к любой другой точке, не выходя из самой фигуры. Вначале может показаться, что это одно и то же, но такое поспешное умозаключение опровергается примером фигуры F.

Формализация этих идей приводит к следующим определениям.

а) Множество $M \subset \mathbb{R}^n$ называется несвязным, если его можно представить в виде объединения двух непересекающихся непустых относительно открытых подмножеств A и B. «Открытых» здесь можно заменить на «замкнутых». Более того, если такое представление возможно, то соответствующее подмножество A является относительно открыто-замкнутым (т. е. и открытым, и замкнутым в M) подмножеством, которое не совпадает ни с $\varnothing$, ни с M; обратно, если существует относительно открыто-замкнутое $A \subset M$, $A \neq \varnothing$, M, то M несвязно. Можно дать формулировку, в которой не прибегают к относительным понятиям: M несвязно, если M можно представить в виде объединения таких двух непустых его подмножеств A и B, каждое из которых не содержит ни одной точки прикосновения другого (см. рис. 25), т. е.

$$M = A \cup B, \quad A \cap \operatorname{clos} B = \varnothing, \quad \operatorname{clos} A \cap B = \varnothing, \qquad A, B \neq \varnothing.$$

Если такое представление невозможно, то пространство называется связным.

Например, отрезок $[a, b]$ связен. Действительно, пусть $[a, b] = A \cup B$, где A, B — непересекающиеся открыто-замкнутые подмножества $[a, b]$ и, скажем, $a \in A$ (последнее ведь зависит просто от выбора обозначений для двух открыто-замкнутых частей $[a, b]$). Докажем, что тогда $A = [a, b]$. Допустим, что $B \neq \varnothing$, и обозначим $c := \inf B$. Будучи замкнутым подмножеством замкнутого отрезка, B является ограниченным замкнутым подмножеством $\mathbb{R}$, и потому $c \in B$. Далее, все точки x, лежащие строго левее c на отрезке $[0, 1]$, должны принадлежать A; а такие точки существуют, ибо из $a \notin B$ следует, что $a \neq c$ и, значит, $a < c$. Итак, $[a, c)$ — непустое подмножество A. Но тогда, ввиду замкнутости A, $c \in A$, и мы приходим к противоречию.

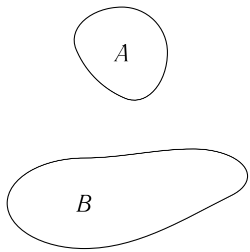


Рис. 25. Несвязность: 2 куска множества $M = A \cup B$

Аналогичным образом доказывается связность открытого интервала (a, b), открытого с одной стороны и замкнутого с другой стороны полуинтервала $[a, b)$ или $(a, b]$, открытой или замкнутой полупрямой $(-\infty, a)$, $(-\infty, a]$, $(a, \infty]$ или $[a, \infty)$, прямой $\mathbb{R}$. Впрочем, формально проще всего вывести их связность из их выпуклости, а в конечном счёте — из связности $[a, b]$; см. ниже.

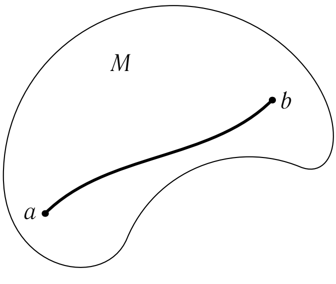


Рис. 26. Путь, соединяющий a и b

б) (Непрерывным) путём в множестве M называется непрерывное отображение $f\colon [a, b] \to M$ отрезка $[a, b]$ в M (в порядке стандартизации обычно берут отрезок $[0, 1]$). Наглядно можно представлять себе, что в M движется точка, занимающая в момент времени $t \in [a, b]$ положение $f(t)$. Точка $x = f(a)$ называется началом пути f, а $y = f(b)$ — его концом; говорят также, что путь f соединяет точки x и y (друг с другом) (см. рис. 26). Множество M называется линейно связным, если любые две его точки можно соединить путём (в M). В противном случае его можно назвать линейно несвязным.

Обратите внимание, что смысл термина «путь» не совсем согласуется со значением этого слова в обычном языке: наш путь — это не только «дорога», по которой движется (с изменением «времени» t) точка $f(t)$, но и описание того, как это движение происходит с учётом времени. Понятие «пути» даёт некоторую формализацию наглядному представлению о «параметризованной кривой». Наглядной интуиции понятие «кривой, т. е. линии» может представляться более привычным, чем понятие «параметризованной кривой» (когда мы вводим на кривой параметр, мы как бы добавляем к простой геометрической картине некий новый и не очень существенный элемент). Но на самом деле при наличии у кривой самопересечений (даже совсем несложных) не так-то просто дать чёткое определение понятию «кривой» (тогда как определение пути очень просто). См. [8], параграф о непрерывных кривых в метрических пространствах[40]. Заметим, что поскольку ничего, кроме непрерывности, мы не требуем, «кривая» может быть, с точки зрения «наивной интуиции», довольно патологической (например, она может быть «кривой Пеано», проходящей через каждую точку квадрата, см. [6] или [7]). Причина, по которой мы можем не уточнять понятия кривой, а говорить о путях, состоит в том, что мы обычно начинаем с параметризованной кривой, т. е. с пути, но затем оказывается, что наши результаты не меняются не только при изменениях параметризации кривой в духе уточнений из [8], но и при гораздо более общих деформациях пути.

С геометрической точки зрения, идея линейной связности состоит в том, что любые две точки $x, y \in M$ можно соединить линией (отсюда и прилагательное «линейная»). Не уточняя понятия «линии», можно согласиться, что, во всяком случае, раз x и y соединены «линией», то от x можно перейти к y, двигаясь вдоль этой линии, причём процесс движения непрерывен. Мы приходим к (тоже достаточно наглядному) «кинематическому» представлению о движущейся точке, которая в момент времени t занимает положение $f(t)$, причём при $t = 0$ она «выходит» из x, а при $t = 1$ «приходит» в y. А это и значит, что существует непрерывное отображение $f\colon [0, 1] \to M$ с $f(0) = x$, $f(1) = y$. Итак, мы можем не вникать в точный смысл понятия кривой, а иметь дело только с простым понятием пути.

Несвязное множество M не может быть линейно связным. Действительно, если $M = A \cup B$ с непересекающимися открытыми A, B, то никакой путь $f\colon [a, b] \to M$ не может соединять точку из A с точкой из B,

[40] Не воспроизводя имеющегося там определения, приведу две фразы оттуда, в которых выражена его идея: «Порядок, в котором проходятся точки кривой, мы будем считать существенным свойством самой кривой.» Этот порядок в [8] вводится с помощью параметризации кривой (при этом точки самопересечения как бы считаются несколько раз). «Однако при одинаковом порядке прохождения точек пространства выбор „параметра“ t мы будем считать несущественным.»

ибо в противном случае множества $f^{-1}(A)$ и $f^{-1}(B)$ были бы непересекающимися открытыми подмножествами отрезка $[a, b]$, вопреки связности последнего.

Стало быть, линейно связное множество является связным. Обратное же не обязательно, как показывает пример подмножества F плоскости $\mathbb{R}^2$.

Очевидна линейная связность (а значит, и связность) любого выпуклого подмножества M евклидова пространства (по определению, любые две точки $x, y \in M$ можно соединить путём $(1-t)x + ty$, $t \in [0, 1]$).

Упражнение 15. Докажите, что непрерывная числовая функция на связном множестве M, принимающая только целочисленные значения, постоянна.

Упражнение 16. Докажите следующие утверждения. Образ связного (или линейно связного) множества при непрерывном отображении связен (линейно связен). Если любые две точки множества M содержатся в некотором связном (линейно связном) подмножестве (так сказать, соединяются таким подмножеством), то M связно (линейно связно). Произведение нескольких связных (линейно связных) множеств связно (линейно связно). Объединение любой системы связных (или линейно связных) подмножеств, имеющих непустое пересечение, связно (линейно связно).

Из последнего утверждения следует, что для любой точки $x \in M$ существует наибольшее связное (линейно связное) подмножество M, содержащее эту точку. Оно называется компонентой связности (линейной связности) точки x (в M). Компонента линейной связности x состоит из всех тех точек, которые можно соединить путём с x.

Упражнение 17. Докажите, что замыкание связного подмножества связно (откуда явствует, что компонента связности любой точки $x \in M$ является замкнутым подмножеством M), и покажите (например, с помощью рассматривавшейся выше фигуры F), что замыкание линейно связного подмножества не обязательно линейно связно.

Упражнение 18. Докажите, что компоненты связности (линейной связности) двух точек множества M либо совпадают, либо не пересекаются.

Отсюда следует, что M является объединением непересекающихся подмножеств — различных компонент связности (линейной связности). В связи с этим их называют также компонентами связности (линейной связности) пространства M или его связными (линейно связными) компонентами.

Когда какое-нибудь множество M (не обязательно топологическое пространство) представлено как объединение своих непустых непересекающихся подмножеств, то, как известно, говорят, что эти подмножества образуют разбиение исходного множества M. Таким образом, мы имеем два разбиения M — на его компоненты связности и на его компоненты линейной связности. Второе из них «мельче» первого, а первое «крупнее» второго в том смысле, что каждый элемент второго разбиения целиком содержится в некотором элементе первого разбиения. Мы видели, что иногда эти два разбиения не совпадают. Всё же они совпадают в наиболее интересном для нас случае.

Упражнение 19. Докажите, что у открытого подмножества $U \subset \mathbb{R}^n$ компоненты линейной связности являются открытыми множествами и что в этом случае наши два разбиения совпадают (так что в данном случае нетренированная интуиция, не различающая связность и линейную связность, не делает ошибки).

Упражнение 20. В [8] открытое подмножество $U \subset \mathbb{R}^n$ названо связным, если любые две его точки x, y можно соединить ломаной, целиком лежащей в U. Докажите, что если точки $x, y \in U$ можно соединить в U путём, то их можно соединить и ломаной, так что определение, данное в [8], совпадает с нашим.

Связные открытые подмножества $\mathbb{R}^n$ называются областями. Однако порой так называют вообще открытые множества (быть может, даже не подмножества $\mathbb{R}^n$), независимо от того, связны они или нет. (И, конечно, слово «область» может использоваться совсем в другом смысле, — например, «область определения функции».)

Пусть A — замкнутое подмножество $\mathbb{R}^n$. Множество $\mathbb{R}^n \setminus A$ — открытое. Говорят, что A разбивает $\mathbb{R}^n$ на области, являющиеся компонентами связности дополнения $\mathbb{R}^n \setminus A$ (это открытое множество), а если компонент всего одна, т. е. если $\mathbb{R}^n \setminus A$ связно, — что A не разбивает $\mathbb{R}^n$.

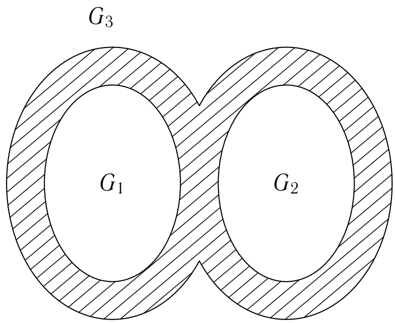


Рис. 27. Заштриховано множество A; G_1, G_2, G_3 — компоненты связности $\mathbb{R}^2 \setminus A$; G_3 — внешняя область

Заметим, что если A ограничено, то одна из компонент связности $\mathbb{R}^n \setminus A$ содержит все точки w с достаточно большими $|w|$, — именно, с $|w| > \max\limits_{z \in A} |z|$. Действительно, дополнение $\{w;\ |w| > r\}$ к кругу $\{w;\ |w| \leqslant r\}$ — связное множество (почему?), которое при $r = \max\limits_{z \in A} |z|$ не пересекается с A и, значит, целиком лежит в одной компоненте связности множества $\mathbb{C} \setminus A$. Эту компоненту называют внешней областью для A или внешней областью множества A (см. рис. 27).

После всего сказанного мы можем уточнить формулировку теоремы Жордана.

Теорема Жордана. *Пусть* Σ^1 — *топологическая окружность* (*т. е. множество, гомеоморфное* $\mathbb{S}^1$) *на плоскости* $\mathbb{C}$. *Тогда дополнение* $\mathbb{C} \setminus \Sigma^1$ *имеет ровно две компоненты связности.*

Мы дали чёткие формальные определения ряда понятий и доказали несколько простых утверждений (нетривиальна только теорема о продолжении). Всё это более или менее согласуется с нашими интуитивными, или, вернее, основанными на повседневном опыте представлениями. Возможно, два вида связности всё-таки вносят некоторую дисгармонию с этими представлениями, но в основном мы скорее уточняли, чем меняли последние. Теорема Жордана, конечно, тоже им отнюдь не противоречит.

Следует предупредить, что на самом деле в области применимости введённых выше понятий имеются несоответствия с доматематическими представлениями, по-своему не менее резкие, чем, скажем, противоречие между наглядным представлением о движении точки, имеющей (кроме, может быть, каких-то исключительных моментов времени) мгновенную скорость, и примером непрерывной нигде не дифференцируемой функции. В простых случаях граница области на плоскости — это одна или несколько замкнутых кривых, но граница может быть и более сложной. Л. Брауэр и Вада указали пример компактного множества K на плоскости, разбивающего её на три области, причём граница каждой из них совпадает со всем K. (Тривиальная модификация их построения приводит к любому другому числу областей с общей границей K, даже бесконечному.) Об этом можно прочитать в [6] или [7]. В своё время соответствующие примеры возникли при продумывании точного смысла тех или иных понятий. Позднее более или менее сходные примеры стали появляться сперва в самой математике при исследовании совсем других (не связанных с анализом смысла понятий) вопросов, а затем и в некоторых вопросах математического естествознания.

§ 4. Отображения окружности в окружность

Пусть $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^m$ и $f_0, f_1 \colon A \to B$ — непрерывные отображения. Гомотопия отображения f_0 в f_1, или гомотопия, соединяющая (связывающая) эти отображения, — это такое семейство $\{f_t\}$ отображений

$$f_t \colon A \to B, \qquad 0 \leqslant t \leqslant 1,$$

что $f_t(x)$ непрерывно зависит от (x, t) (т. е. непрерывно отображение

$$F \colon A \times [0, 1] \to B, \qquad F(x, t) = f_t(x),$$

которое тоже называют гомотопией), а $f_0 = f$ и $f_1 = g$. Если существует гомотопия, соединяющая f и g, то говорят, что f и g гомотопны (друг другу) и пишут $f \sim g$.

В этом случае говорят также, что отображение f деформируется в отображение g и что семейство отображений $\{f_t\}$ есть соответствующая деформация. Это отвечает наглядному представлению: точки $f_t(x)$ движутся с изменением «параметра деформации» (или «параметра гомотопии») t, играющего в этой картине движения роль «времени». Обращаю внимание, что здесь употребление слова «деформация» отличается от его употребления в § 1 при первом наглядном пояснении предмета топологии. Во-первых, там речь шла о деформациях фигуры, а не отображения, а во-вторых (и это более существенно), там молчаливо подразумевалось, что у деформируемой фигуры всё-таки не возникает самопересечений (что соответствует наглядному представлению о деформации «резиновой» фигуры). При гомотопии же вполне может рассматриваться даже такое отображение $f\colon A \to B$, при котором всё A отображается в одну точку («отображение в точку»). Кстати, непрерывные отображения, гомотопные отображению в точку, играют роль, напоминающую роль нуля. Поэтому их часто называют «гомотопными нулю». Их называют также «несущественными отображениями» [41].

Очевидно, что $f \sim f$.

Упражнение 21. Докажите, что если $f \sim g$, то $g \sim f$. (Указание: рассмотрите $\{f_{1-t}\}$.)

Пусть имеются непрерывные отображения $f, g, h\colon A \to B$, причём $f \sim g$ и $g \sim h$. Тогда $f \sim h$. Очевидная идея доказательства состоит в том, чтобы сперва продеформировать f в g, а затем, продолжая процесс деформации, продеформировать g в h. Пусть $\{f_t\}$ — гомотопия f в g, а $\{g_t\}$ — гомотопия g в h. В обоих этих семействах отображений параметр деформации t принимает значения от 0 до 1. Изменим параметр деформации во втором семействе так, чтобы он принимал значения от 1 до 2; для этого надо рассматривать второе семейство как $\{g_{t-1};\ 1 \leqslant t \leqslant 2\}$. Теперь можно соединить эти два семейства отображений в одно:

$$h_t\colon A \to B, \qquad h_t(x) = \begin{cases} f_t(x) & \text{при } 0 \leqslant t \leqslant 1, \\ g_{t-1}(x) & \text{при } 1 \leqslant t \leqslant 2. \end{cases}$$

В терминах соответствующих отображений F, G, мы «склеиваем» два непрерывных отображения, полагая

$$H\colon A \to B, \qquad h(x,\, t) = \begin{cases} F(x,\, t) & \text{при } 0 \leqslant t \leqslant 1, \\ G(x,\, t-1) & \text{при } 1 \leqslant t \leqslant 2. \end{cases}$$

[41] Я буду время от времени пользоваться последним термином, но должен предупредить, что он является не совсем общепринятым. По крайней мере, в [2] под несущественными отображениями $A \to \mathbb{S}^1$ понимают отображения несколько иного типа. (Правда, для компактных A определение из [2] оказывается эквивалентным нашему. Но нам понадобятся и некомпактные A.)

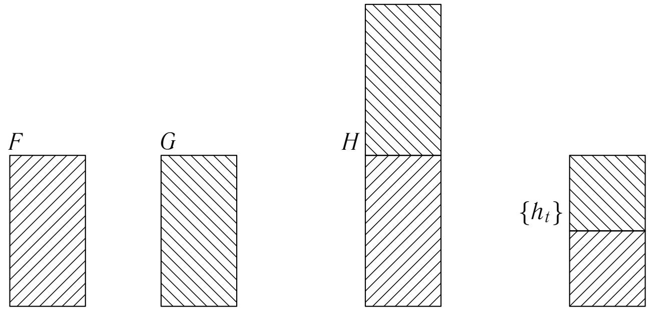


Рис. 28. Прямоугольники условно изображают $A \times [0, 1]$ или $A \times [0, 2]$. Штриховки означают, что на прямоугольнике задано отображение в B.

(Условно это показано на рис. 28.) H корректно определено и непрерывно: при $t = 1$ (где пересекаются «области действия» обеих строчек) обе строчки предписывают одно и то же (а именно, что $H(x, 1) = g(x)$); эти «области действия» суть замкнутые множества $A \times [0, 1]$ и $A \times [1, 2]$). Только теперь $t \in [0, 2]$, а в определении гомотопии в порядке стандартизации $t \in [0, 1]$. Это, конечно, несущественно: надо ввести новый параметр деформации, равный $t/2$, так что окончательно мы чуть изменяем определение $\{h_t\}$:

$$h_t\colon A \to B, \qquad h_t(x) = \begin{cases} f_{2t}(x) & \text{при } 0 \leqslant t \leqslant 1/2, \\ g_{2t-1}(x) & \text{при } 1/2 \leqslant t \leqslant 1. \end{cases}$$

Последние три свойства гомотопии выражают словами: гомотопия является отношением эквивалентности. Если читатель не знаком с последним понятием, то это неважно [42]. Просто я сказал, что последние три свойства похожи на известные свойства равенства, и хотя гомотопия — всё-таки отнюдь не равенство, эти её свойства позволяют в известных случаях заменять гомотопные отображения друг другом. А вот ещё одно простое свойство гомотопии, в котором говорится о её поведении при композиции отображений [43] и которое тоже позволяет порой заменять гомотопные отображения друг другом.

Пусть A, B, C — подмножества каких-то $\mathbb{R}^n$ и пусть имеются непрерывные отображения $f_0, f_1\colon A \to B$ и $g_0, g_1\colon B \to C$, причём

[42] А если хочет с ним познакомиться или освежить соответствующие сведения в памяти, то может обратиться к [8].

[43] Пусть имеются два отображения каких-нибудь множеств $f\colon A \to B$, $g\colon B \to C$. Их композицией (или суперпозицией) $g \circ f$ называют отображение $A \to C$, при котором элемент $x \in A$ переходит в $g(f(x))$ (более наглядно: $x \xrightarrow{f} f(x) \xrightarrow{g} g(f(x))$).

$f_0 \sim f_1$ *и* $g_0 \sim g_1$. *Тогда* $g_0 \circ f_0 \sim g_1 \circ f_1$. Если $\{f_t\}$ и $\{g_t\}$ суть соответствующие гомотопии, то в качестве искомой гомотопии можно взять $\{g_t \circ f_t\}$. В терминах соответствующих отображений $F\colon A \times I \to B$, $G\colon B \times I \to C$ речь идёт об отображении

$$H\colon A \times I \to C, \qquad H(x,\, t) = G(F(x,\, t),\, t),$$

откуда очевидна непрерывность построенной гомотопии по $(x,\, t)$.

На понятии гомотопии основан важнейший раздел топологии — гомотопическая топология. Грубо говоря, для неё характерна точка зрения «с точностью до гомотопии». С этой точки зрения гомотопные отображения не различаются по своим свойствам. Не различаются также гомотопические свойства фигур A и B, если существуют непрерывные отображения $f\colon A \to B$ и $g\colon B \to A$, которые с гомотопической точки зрения являются как бы взаимно обратными; это значит, что если перейти из A в B с помощью f, а затем вернуться в A с помощью g, то полученное отображение $g \circ f\colon A \to A$ будет гомотопно тождественному отображению $1_A\colon A \to A$, оставляющему все точки A на месте, и что отображение $f \circ g\colon B \to B$ гомотопно тождественному отображению $1_B\colon B \to B$. В этом случае говорят, что A и B гомотопически эквивалентны.

По сравнению с точкой зрения «с точностью до гомеоморфизма» точка зрения «с точностью до гомотопической эквивалентности» является значительно более «грубой» и «примитивной». Может показаться, что такой подход является уж слишком грубым. Но это не так. Во-первых, оказывается, что всё-таки не все фигуры гомотопически эквивалентны друг другу. Например, можно доказать, что n-мерные сферы

$$\mathbb{S}^n = \{x \in \mathbb{R}^{n+1};\ |x| = 1\}$$

при различных n гомотопически не эквивалентны. Во-вторых, бывает, что задачи, вовсе не связанные с таким грубым подходом, удаётся свести к гомотопическим задачам, а последние решить. Например, хотя «чистая» гомотопическая топология «не чувствует» различие между пространствами $\mathbb{R}^n$ и $\mathbb{R}^m$ с $n \neq m$, которые оказываются гомотопически эквивалентными, после небольшого дополнительного построения приходим к задаче, где различие уже чувствуется. В середине XX в. прогресс топологии был в значительной степени связан с гомотопической топологией, — и сама она развивалась, и изобретались приёмы, сводившие задачи не-гомотопического характера к гомотопическим.

Упражнение 22. а) Докажите, что любое непрерывное отображение $f\colon A \to \mathbb{R}^n$ гомотопно нулю. (Указание. $f_t(x) = (1 - t)f(x)$, см. рис. 29.) Выведите отсюда, что $\mathbb{R}^n$ гомотопически эквивалентно фигуре, состоящей из одной-единственной точки, а потому пространства $\mathbb{R}^n$ с различными n гомотопически эквивалентны друг другу.

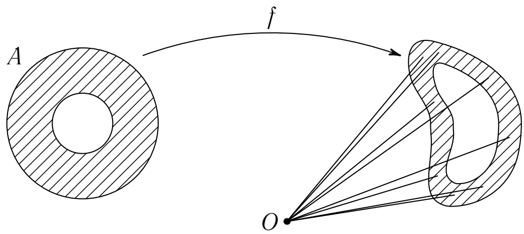


Рис. 29

б) Докажите, что $\mathbb{R}^n \setminus \{0\}$ и $\mathbb{S}^{n-1}$ гомотопически эквивалентны. (Указание: радиусы, см. рис. 30.) Выведите отсюда, что если $\mathbb{R}^n$ и $\mathbb{R}^m$ гомеоморфны, то $\mathbb{S}^{n-1}$ и $\mathbb{S}^{m-1}$ гомотопически эквивалентны (а из сказанного выше явствует, что тогда $n = m$).

Таким образом, опираясь в основном на гомотопическую топологию и привлекая всего лишь совсем простое дополнительное построение, мы смогли доказать, что $\mathbb{R}^n$ и $\mathbb{R}^m$ не гомеоморфны друг другу. Следует заметить, что имеется взаимно однозначное соответствие $\mathbb{R}^n \leftrightarrow \mathbb{R}^m$, которое не непрерывно, и что даже при $n < m$ имеется непрерывное отображение $\mathbb{R}^n$ на всё $\mathbb{R}^m$ (с точностью до несущественных деталей это уже упоминавшаяся «кривая Пеано»). Поэтому доказать, что при различных n пространства $\mathbb{R}^n$ не гомеоморфны друг другу в своё время было серьёзной проблемой, решение которой было впервые получено совсем иначе. Мы видим, что гомотопическая топология легко справляется с этой задачей. Правда, в настоящей книжке полное доказательство отсутствует — в ней не доказывается, что $\mathbb{S}^n$ при различных n гомотопически не эквивалентны. Последнее делается не голыми руками, но прошу поверить, что в гомотопической топологии такой результат получается на довольно раннем этапе.

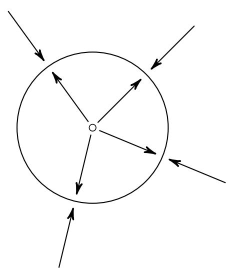


Рис. 30

Доказательство теоремы Жордана, которое мы дадим, является другим примером применения гомотопических соображений к задаче не-гомотопического характера.

Перейдём от пропаганды к делу. Ниже (как это обычно в топологии) отрезок $[0, 1]$ обозначается через I.

Пусть $A \subset \mathbb{R}^n$. Угловая функция непрерывного отображения $f\colon A \to \mathbb{S}^1$ — это такая непрерывная функция $\varphi\colon A \to \mathbb{R}$, что $f(x) = e^{i\varphi(x)}$ при всех $x \in A$. Она не всегда существует. Ведь если у отображения f имеется угловая функция φ, то $\{f_t(x) = e^{it\varphi(x)};\ t \in I\}$ — гомотопия, соединяющая два отображения: отображение в точку (с образом 1), получающееся при

$t = 0$, и исходное отображение f, получающееся при $t = 1$; стало быть, f несущественно. Однако для некоторых фигур, как мы увидим, имеются и такие непрерывные отображения $f\colon A \to \mathbb{S}^1$, которые не гомотопны нулю. Ниже мы увидим, что и обратно, отображение, гомотопное отображению в точку, имеет угловую функцию.

Угловая функция (если она существует) не единственна: при изменении её на кратное 2π снова получится угловая функция (ведь $e^{2\pi i} = 1$). Если A связно и φ, ψ — две угловые функции непрерывного отображения $f\colon A \to \mathbb{S}^1$, то φ и ψ различаются на (постоянное) кратное 2π (в частности, если они совпадают в какой-то точке, то совпадают всюду). Действительно, $e^{i(\varphi(x)-\psi(x))} = 1$ при всех x, так что непрерывная функция $\varphi(x)-\psi(x)$ при каждом x является кратным 2π, а $\dfrac{1}{2\pi}(\varphi(x) - \psi(x))$ — непрерывная функция с целочисленными значениями. Ввиду связности A такая функция постоянна (упражнение 15).

Теорема. *Непрерывное отображение $f\colon I \to \mathbb{S}^1$ имеет угловую функцию φ, для которой $\varphi(0) = \alpha$ — заранее выбранный аргумент числа $f(0)$.*

Идея доказательства состоит в том, что а) образ $f([0, s])$ маленького отрезка $[0, s]$ целиком содержится в небольшой дуге возле точки $e^{i\alpha}$, откуда, как мы увидим, легко вывести существование у f угловой функции на этом отрезке, близкой к числу α; и б) если мы каким-то образом построили угловую функцию на отрезке $[0, s]$, то эту функцию можно продолжить на немного бóльший отрезок $[0, s']$, потому что образ $f([s, s'])$ отрезка $[s, s']$ целиком содержится в небольшой дуге возле точки $f(s) = e^{i\varphi_s}$, так что для $t \in [s, s']$ можно определить φ_t путём небольшого изменения φ_s. Если в а) $\alpha = 0$, то на маленьком отрезке $[0, s]$ роль угловой функции играет просто $\widehat{\arg} f(s)$; случай же $\alpha \neq 0$ легко сводится к предыдущему — мы строим угловую функцию для «подкрученного» отображения $s \mapsto e^{-i\alpha} f(s)$ (для него α заменяется на 0 — образ $[0, s]$ близок к 1) и добавляем к ней α (компенсируя «подкрутку»). Аналогично осуществляется в б) продолжение угловой функции на $[s, s']$. По существу, таким путём получается «конструктивное» доказательство нашей теоремы, при котором φ_s последовательно определяется на некоторых отрезках $[0, s_i]$ всё бóльшей и бóльшей длины, и за конечное число шагов отрезок $[0, s_i]$ сравняется со всем отрезком $[0, 1]$. Однако аккуратно доказательство проще оформить несколько иначе; внешне оно не будет выглядеть конструктивным.

Доказательство. Рассмотрим множество

$$A = \{t;\ \text{у отображения } f|[0, s]\colon [0, s] \to \mathbb{S}^1 \text{ имеется угловая функция } \varphi_s \text{ с } \varphi_s(0) = \alpha.\}$$

Нам надо доказать, что $1 \in A$.

Содержит ли A число 0 — это условность: считаем ли мы возможным говорить об «отрезке» $[0, 0] = \{0\}$ и о соответствующей угловой функции, сводящейся к тому, что нулю сопоставлено α? Но уже без всяких условностей A содержит близкие к 0 положительные числа. Действительно, если $z > 0$ достаточно мало, то $f([0, s]) \not\ni \{-1\}$. А тогда φ_s определяется как $\widehat{\arg}\dfrac{f(t)}{f(0)} + \alpha$. (Непрерывность явствует из того, что $f(t) \notin (-\infty, 0]$ ни при одном $t \in [0, s]$; ясно также, что $e^{i\varphi_s(t)} = \dfrac{f(t)}{f(0)}e^{i\alpha} = f(t)$.)

Пусть $a = \sup A$. По сказанному, $a > 0$. Для $t \in I$, достаточно близких к a, — скажем, при $t \in [a - \delta, a + \delta] \cap I$ — будет $\operatorname{Re}\dfrac{f(t)}{f(a)} > 0$ (ведь при $t = a$ это 1). Некоторое $s \in A$ попадает в $[a - \delta, a]$, — пусть это будет s_0. Поскольку при $t \in [a - \delta, a + \delta] \cap I$

$$\left|\widehat{\arg}\frac{f(t)}{f(a)}\right| < \frac{\pi}{2} \qquad \text{и} \qquad \left|\widehat{\arg}\frac{f(a)}{f(s_0)}\right| = \left|-\widehat{\arg}\frac{f(s_0)}{f(a)}\right| < \frac{\pi}{2},$$

то $\dfrac{f(t)}{f(s_0)} \neq -1$ при всех таких t. Положим $\psi(t) = \widehat{\arg}\dfrac{f(t)}{f(s_0)} + \varphi_{s_0}(s_0)$. Это непрерывная функция на $[a - \delta, a + \delta] \cap I$ и

$$e^{i\psi(t)} = e^{i\varphi_{s_0}(t)}\frac{f(t)}{f(s_0)} = f(s_0)\frac{f(t)}{f(s_0)} = f(t).$$

Кроме того, $\psi(s_0) = \varphi_{s_0}(s_0)$. Поэтому и $\varphi_{s_0}|[a - \delta, s_0]$, и $\psi|[a - \delta, s_0]$ являются угловыми функциями отображения $f|[a - \delta, s_0]$, а при $s = s_0$ они совпадают; значит, они совпадают всюду на $[a - \delta, s_0]$. Поэтому функция

$$\varphi(s) = \begin{cases} \varphi_{s_0}(s) & \text{при } s \in [0, s_0], \\ \psi(s) & \text{при } s \in [a - \delta, a + \delta] \cap I \end{cases}$$

корректно определена и непрерывна. Ясно, что $\varphi(0) = \alpha$ и что $f(t) = e^{i\varphi(t)}$ при $t \in I \cap [0, a + \delta]$. В частности, если бы было $a < 1$, то A содержало бы число $a' = \min(1, a + \delta/2)$, которое больше a, хотя $a = \sup A$. А если $a = 1$, то область определения φ как раз и есть $[0, 1] = I$.

Теорема. *Если $f\colon A \to \mathbb{S}^1$ несущественно, то у f имеется угловая функция.*

Доказательство. В данном случае несущественное отображение f гомотопно отображению f_0, переводящему всё A в точку 1, — ведь если f гомотопно отображению в какую-то другую точку, то эту точку можно непрерывно перевести в 1 (уточните это соображение!). Пусть $\{f_t\}$ — гомотопия, соединяющая отображение f_0 с $f_1 = f$ (в данном случае нам удобно параметризовать гомотопию именно таким образом). При каждом $x \in A$ рассмотрим путь, по которому движется образ этой точки при гомотопии, т. е. отображение

$$I \to \mathbb{S}^1, \qquad t \mapsto f_t(x).$$

Оно имеет угловую функцию $\varphi_t(x)$, принимающую при $t = 0$ значение 0 (это один из аргументов числа 1). Читатель, конечно, уже догадался, что

$$A \times I \to \mathbb{R}, \qquad (x,\, t) \mapsto \varphi_t(x)$$

является угловой функцией отображения

$$A \times I \to \mathbb{S}^1, \qquad (x,\, t) \mapsto f_t(x).$$

Но здесь кое-что ещё надо доказать: сказав, что $t \mapsto \varphi_t(x)$ является угловой функцией некоторого отображения $I \to \mathbb{S}^1$, мы говорим тем самым, что $\varphi_t(x)$ непрерывна по t при фиксированном x, нам же нужна непрерывность $\varphi_t(x)$ по $(x,\, t)$.

У каждой точки $x_0 \in A$ имеется такая окрестность U в A, что

$$\frac{f_t(x)}{f_t(x_0)} \neq -1 \quad \text{при} \quad (x,\, t) \in U \times I. \tag{28}$$

Действительно, если бы это было не так, то существовала бы такая последовательность $(x_j,\, t_j)$, что $x_j \to x_0$ и $\dfrac{f_{t_j}(x_j)}{f_{t_j}(x_0)} = -1$. Выберем из $\{t_j\}$ сходящуюся подпоследовательность $\{t_{j_k}\}$; пусть $t_0 = \lim t_{j_k}$. Тогда предельный переход в равенстве $f_{t_{j_k}}(x_{j_k}) = -f_{t_{j_k}}(x_0)$ приводит к выводу, что $f_{t_0}(x_0) = -f_{t_0}(x_0)$, а это было бы возможно только при $f_{t_0}(x_0) = 0$, тогда как все $|f_t(x)| = 1$.

Пусть U — такая окрестность точки x_0 в A, для которой справедливо (28). При $(x,\, t) \in U \times I$ положим

$$\psi_t(x) = \widehat{\arg}\, \frac{f_t(x)}{f_t(x_0)} + \varphi_t(x_0).$$

В своей области определения $\psi_t(x)$ является непрерывной функцией от $(x,\, t)$, причём

$$e^{i\psi_t(x)} = \frac{f_t(x)}{f_t(x_0)} e^{i\varphi_t(x_0)} = \frac{f_t(x)}{f_t(x_0)} f_t(x_0) = f_t(x).$$

В частности, при фиксированном $x \in U$ функция

$$I \to \mathbb{R}, \qquad t \mapsto \psi_t(x)$$

является угловой функцией отображения

$$I \to \mathbb{S}^1, \qquad t \to f_t(x).$$

При $t = 0$ она принимает значение

$$\psi_0(x) = \widehat{\arg}\,\frac{f_0(x)}{f_0(x_0)} + \varphi_0(x_0) = \widehat{\arg}\,1 + 0 = 0,$$

т. е. то же самое значение, что и угловая функция $\varphi_t(x)$ того же отображения. Значит,

$$\varphi_t(x) = \psi_t(x) \quad \text{при всех} \;\; (x,\, t) \in U \times I.$$

Отсюда следует непрерывность $\varphi_t(x)$ по $(x,\, t)$ в точке $(x_0,\, t_0)$ с любым t_0.

А теперь ясно, что функция $\varphi_1(x)$ служит угловой функцией для $f_1 = f$, и теорема доказана.

В частности, любое непрерывное отображение $f\colon \mathbb{R} \to \mathbb{S}^1$ гомотопно отображению в точку (гомотопия $f_t(x) = f(tx)$ даёт f при $t = 1$ и отображение, переводящее всё $\mathbb{R}$ в $f(0)$, при $t = 0$). Поэтому у него всегда имеется угловая функция.

Заметим, что заодно мы доказали следующее утверждение: *пусть*

$$f_t\colon A \to \mathbb{S}^1 \qquad (t \in I) \tag{29}$$

— гомотопия, начинающаяся (при $t = 0$) с отображения f_0, переводящего всё A в точку 1. Тогда у отображений f_t имеются такие угловые функции ψ_t, что $\psi_0(x) = 0$ (при всех $x \in A$) и $\psi_t(x)$ непрерывно по $(x,\, t)$.

Это утверждение можно немного обобщить.

Пусть (29) — гомотопия, начинающаяся (при $t = 0$) с отображения $f_0 = f$, которое имеет угловую функцию φ. Тогда у отображений f_t имеются такие угловые функции ψ_t, что $\psi_0(x) = \varphi(x)$ (при всех $x \in A$) и $\psi_t(x)$ непрерывно по $(x,\, t)$.

Для доказательства рассмотрим семейство отображений

$$g_t\colon A \to \mathbb{S}^1, \qquad g_t(x) = \frac{f_t(x)}{f(x)}.$$

Это гомотопия, удовлетворяющая условиям предыдущего утверждения (проверьте!), поэтому g_t имеют соответствующие угловые функции χ_t ($\chi_0 = 0$ и $\chi_t(x)$ непрерывно по $(x,\, t)$). Тогда $\psi_t(x) = \chi_t(x) + \varphi(x)$ служат угловыми функциями для f_t, имеющими требуемые свойства.

Степень отображения

Отныне мы несколько изменим терминологию для непрерывных отображений $f\colon \mathbb{S}^1 \to \mathbb{S}^1$. Для такого отображения под его угловой функцией мы будем понимать угловую функцию $\psi(\varphi)$ отображения-композиции

$$\mathbb{R} \to \mathbb{S}^1 \to \mathbb{S}^1, \qquad \varphi \mapsto e^{i\varphi} \mapsto f(e^{i\varphi}).$$

Это теперь функция, определённая на $\mathbb{R}$, а не на $\mathbb{S}^1$, как раньше, и она существует для всех непрерывных отображений $\mathbb{S}^1 \to \mathbb{S}^1$, а не только для некоторых. Тот случай, когда f имеет угловую функцию в прежнем смысле, — обозначим её через χ, — в новых терминах характеризуется тем, что $\psi(\varphi) = \chi(e^{i\varphi})$ является периодической функцией от φ с периодом 2π. Как мы видели, такие угловые функции получаются в том и только том случае, когда f несущественно. (С другой стороны, тождественное отображение $\mathbb{S}^1$ на себя, переводящее каждую точку в неё саму, имеет угловую функцию $\psi(\varphi) = \varphi$, которая, очевидно, не периодична. В прежних терминах у него не было угловой функции.)

Угловая функция отображения $\mathbb{S}^1 \to \mathbb{S}^1$ определена не единственным образом: если $\psi(\varphi)$ — одна из угловых функций, то остальные имеют вид $\psi(\varphi) + 2\pi j$ со всевозможными целыми j.

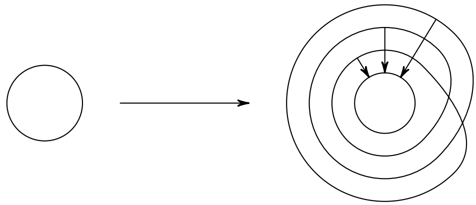


Рис. 31. Окружность наматывается 3 раза на себя

Если $\psi(\varphi)$ — угловая функция для f, то $\psi(\varphi + 2\pi)$ — тоже угловая функция (почему?). Значит,

$$\psi(\varphi + 2\pi) = \psi(\varphi) + 2\pi k \quad \text{с некоторым целым } k. \tag{30}$$

Если вместо угловой функции $\psi(\varphi)$ исходить из угловой функции $\psi_1(\varphi) = \psi(\varphi) + 2\pi j$, то получится, что $\psi_1(\varphi + 2\pi) - \psi_1(\varphi) = 2\pi k$ с тем же самым k. Значит, целое число k не зависит от конкретного выбора угловой функции, а определяется исходным отображением f. Это число называется степенью отображения f. Мы будем обозначать степень через $\text{Ст}\, f$. Она имеет простой наглядный смысл: при непрерывном отображении $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ степени k окружность $\mathbb{S}^1$ наматывается на себя k раз (рис. 31).

Заметим ещё, что любая непрерывная функция, обладающая свойством (30), является угловой функцией некоторого непрерывного отображения $f\colon \mathbb{S}^1 \to \mathbb{S}^1$. Оно определяется по формуле

$$f(e^{i\varphi}) = e^{i\psi(\varphi)}.$$

Надо убедиться в корректности такого определения: ведь точка $\mathbb{S}^1$ имеет бесконечное число аргументов φ. Но все они различаются на кратные 2π, а из (30) видно, что соответствующие $\psi(\varphi)$ тоже различаются на кратные 2π.

Название «степень» связано с простейшим примером — отображением $z \mapsto z^k$, рассматриваемом не при всех $z \in \mathbb{C}$, а только на $\mathbb{S}^1$. Если $z = e^{i\varphi}$, то $z^k = e^{ik\varphi}$, так что одной из его угловых функций является $\psi(\varphi) = k\varphi$. При увеличении аргумента на 2π к $\psi(\varphi)$ добавляется $2\pi k$ — степень в нашем смысле совпадает с той степенью, в которую возводится z.

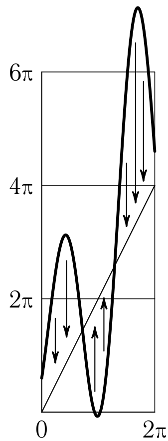


Рис. 32

Отображение f степени k гомотопно отображению $z \mapsto z^k$ (рассматриваемому только на $\mathbb{S}^1$) (см. рис. 32). Действительно, пусть $\psi(\varphi)$ — угловая функция для f. Тогда $\psi(\varphi + 2\pi) - \psi(\varphi) = 2\pi k$. Положим $\psi_t(\varphi) = (1 - t)\psi(\varphi) + tk\varphi$. Тогда $\psi_t(\varphi + 2\pi) = \psi_t(\varphi) + 2\pi k$ при всех $(\varphi, t) \in \mathbb{R} \times I$, так что корректно определены отображения

$$f_t : \mathbb{S}^1 \to \mathbb{S}^1, \qquad f_t(e^{i\varphi}) = e^{i\psi_t(\varphi)},$$

причём $f_t(z)$ непрерывно зависит от (z, t). Доказывая непрерывность в точке (z_0, t_0), где z_0 имеет одним из своих аргументов φ_0, обозначим через U окрестность точки z_0 на $\mathbb{S}^1$, где $\dfrac{z}{z_0} \neq -1$ (на самом деле эта «окрестность» — почти вся $\mathbb{S}^1$). Тогда в U одним из аргументов z является непрерывно зависящее от z число $\varphi(z) = \widehat{\arg}\,\dfrac{z}{z_0} + \varphi_0$. В $U \times I$

$$f_t(z) = e^{i[(1-t)\psi(\varphi(z)) + tk\varphi(z)]},$$

откуда следует непрерывность $f_t(z)$ по (z, t) в $U \times I$ и, в частности, в точке (z_0, t_0). Семейство $\{f_t\}$ — гомотопия, соединяющая отображение f (оно отвечает $t = 0$) с отображением $z \mapsto z^k$ (оно получается при $t = 1$).

Отсюда видно, что *если у непрерывных отображений $f, g : \mathbb{S}^1 \to \mathbb{S}^1$ степени совпадают, то $f \sim g$*. Обратное тоже верно: *пусть $\{f_t\}$ — произвольная гомотопия $\mathbb{S}^1 \to \mathbb{S}^1$. Тогда степень у всех f_t одна и та же.* Действительно, у f_t имеются такие угловые функции ψ_t, что $\psi_t(x)$ непрерывно по (x, t). Тогда

$$\operatorname{Ст} f_t = \frac{1}{2\pi}(\psi_t(\varphi + 2\pi) - \psi_t(\varphi))$$

(φ здесь можно взять любым). Функция

$$\varphi \mapsto \frac{1}{2\pi}(\psi_t(\varphi + 2\pi) - \psi_t(\varphi))$$

— это непрерывная функция на связном множестве I, принимающая целые значения. Такая функция постоянна.

Упражнение 23. Пусть $f, g\colon \mathbb{S}^1 \to \mathbb{S}^1$ — непрерывные отображения. Докажите, что $\text{Ст}(f \circ g) = \text{Ст}\, f \cdot \text{Ст}\, g$.

До сих пор речь шла об отображениях стандартной окружности $\mathbb{S}^1$ в себя. Перейдём к непрерывным отображениям топологических окружностей друг в друга. Пусть топологические окружности S^1 (для неё, в отличие от стандартной окружности, использован обычный шрифт) и Σ^1 являются образами стандартной окружности при гомеоморфизмах $g\colon \mathbb{S}^1 \to S^1$, $h\colon \mathbb{S}^1 \to \Sigma^1$ (тем самым точки S^1 и Σ^1 «параметризуются» числами из $\mathbb{S}^1$), а $f\colon S^1 \to \Sigma^1$ — непрерывное отображение. Под $\text{Ст}\, f$ естественно понимать $\text{Ст}(h^{-1} \circ f \circ g)$, потому что отображение $h^{-1} \circ f \circ g\colon \mathbb{S}^1 \to \mathbb{S}^1$ описывает отображение f в терминах этих параметризаций (если z — «параметр», отвечающий точке $x = g(z) \in S^1$, $f(x) = y$ и w — «параметр», отвечающий точке $y = h(w) \in \Sigma^1$, то мы описываем $x \mapsto y$ как $z \mapsto w$). Но число $\text{Ст}(h^{-1} \circ f \circ g)$ отчасти зависит от конкретного выбора g и h (т. е. от используемых «параметризаций» топологических окружностей S^1 и Σ^1). Уточним характер этой зависимости, для чего понадобится говорить об ориентациях.

Ориентировать топологическую окружность Σ^1 — это значит выбрать на ней положительное направление (тогда говорят об ориентированной топологической окружности). Гомеоморфизм $g\colon \mathbb{S}^1 \to \Sigma^1$ определяет некоторую ориентацию на Σ^1: положительным считается то направление, в котором движется точка $g(e^{i\varphi})$ при возрастании параметра φ (когда, стало быть, $e^{i\varphi}$ движется в положительном направлении по стандартной окружности $\mathbb{S}^1$). Наглядно это понятно, но формальное определение понятия «положительное направление на замкнутой кривой», с какой бы стороны к нему ни подходить, получается довольно громоздким[44]. Следующий подход кажется мне самым коротким, но он может произвести впечатление несколько формального. Сказать, что два гомеоморфизма $\mathbb{S}^1 \to \mathbb{S}^1$

[44] Понятие положительного направления на простой дуге (гомеоморфном образе отрезка) J является более простым. Наглядно понятно, что если такое направление выбрано, то оно позволяет говорить, что одни точки предшествуют другим. Обратно, если мы предпишем, что одни точки предшествуют другим и что это «предшествование» удовлетворяет некоторым довольно простым условиям, то тем самым определяется положительное направление на J. На основании сказанного можно дать определение «ориентации на J» как некоторого указания, когда точка x предшествует точке y. (И всё же авторы [8] сочли даже это слишком громоздким, а предпочли использовать параметризации.) На замкнутой же ориентированной кривой Σ^1 не приходится говорить, что «x предшествует y», ибо, выйдя из x и двигаясь в положительном направлении, мы придём в y, а затем снова в x. Поэтому положительное направление на Σ^1 можно описывать в терминах свойства троек точек x, y, z, наглядно означающего, что если выйти из x и двигаться в положительном направлении, то сперва нам встретится y, а потом z. Уже одно то, что в формулировке надо говорить о тройках x, y, z а не о парах x, y (как в свойстве «x предшествует y» на простой дуге), делает её громоздкой. (Возможна более короткая формулировка в духе [8], связанная с использованием «параметризаций», только на сей раз это будут «параметризации» числами из $\mathbb{S}^1$ или что-то аналогичное.)

одинаково ориентируют $\mathbb{S}^1$ — значит сказать просто то, что они гомотопны. При этом мы оставляем в стороне вопрос, какой геометрический смысл придаётся словам «ориентация Σ^1»; если угодно, можно считать, что так называется класс гомотопных друг другу гомеоморфизмов $\mathbb{S}^1 \to S^1$. Это, во всяком случае, не противоречит геометрической интуиции, хотя едва ли выражает её самым непосредственным образом.

Упражнение 24. Докажите следующие утверждения.

а) Гомеоморфизмы $f, g\colon \mathbb{S}^1 \to \Sigma^1$ гомотопны тогда и только тогда, когда $f^{-1} \circ g \sim 1_{\mathbb{S}^1}$ или, что то же самое, $g^{-1} \circ f \sim 1_{\mathbb{S}^1}$.

б) Имеется ровно два класса гомотопных гомеоморфизмов $\mathbb{S}^1 \to \Sigma^1$. При переходе от одного из них к другому говорят об изменении ориентации Σ^1 на противоположную.

в) Определение степени отображения топологических окружностей $f\colon S^1 \to \Sigma^1$ как $\mathrm{Ст}(h^{-1} \circ f \circ g)$ относится к ориентированным окружностям (ориентации доставляются гомеоморфизмами g и h — при указанном выше понимании слов «ориентация топологической окружности» это суть просто классы гомеоморфизмов $\mathbb{S}^1 \to S^1$ и $\mathbb{S}^1 \to \Sigma^1$, содержащие, соответственно, g и h). При изменении одной из ориентаций у $\mathrm{Ст}\, f$ изменяется знак. (Указание: $h_1^{-1} \circ f \circ g_1 = (h_1^{-1} \circ h) \circ (h^{-1} \circ f \circ g) \circ (g^{-1} \circ g_1)$.)

Если $\Sigma^1 \subset \mathbb{C}$, то хотелось бы указать, какую из двух возможных ориентаций Σ^1 можно считать связанной с той ориентацией, которая имеется в $\mathbb{C}$. Идея здесь могла бы быть той же, как для $\mathbb{S}^1$: мы берём точку O, расположенную внутри Σ^1, принимаем её за полюс и смотрим, насколько изменится полярный угол точки A, лежащей на Σ^1, когда эта точка делает один обход по Σ^1. Если угол изменяется на 2π, то направление обхода считается положительным, а если на -2π, то отрицательным. Но здесь необходимо выяснить два вопроса.

1. Говоря о точке, расположенной внутри Σ^1, мы, конечно, подразумеваем, что эта точка принадлежит той из двух компонент связности $\mathbb{C} \setminus \Sigma^1$, которая не является внешней (эту компоненту естественно назвать внутренней). Здесь мы опираемся на теорему Жордана, которая у нас пока не доказана.

2. Надо проверить, что при обходе A по Σ^1 изменение полярного угла при любом выборе полюса во внутренней области одинаково и равно 2π или -2π. Это формально не следует из формулировки теоремы Жордана, но будет установлено в ходе её доказательства.

Ниже нам придётся говорить о степени некоторого отображения $\Sigma^1 \to \mathbb{S}^1$. Ориентация Σ^1 при этом будет определяться выбором гомеоморфизма $\mathbb{S}^1 \to \Sigma^1$ («параметризации» Σ^1). Нам, к счастью, не понадобится связывать её с ориентацией $\mathbb{C}$. Это, конечно, вносит элемент произвола, но он не влияет на результаты. (Ситуация такая же, как и тогда, когда мы используем координаты, в выборе которых имеется известный произвол, при доказательстве геометрической теоремы, в которой эти координаты не упоминаются.)

Связь степени отображения окружности с прообразами точки

В настоящей книжке материал данного пункта используется только во второй половине § 6. Но на самом деле он весьма важен. Достаточно сказать, что для отображений сфер (и вообще для так называемых многообразий — эти объекты включают все замкнутые поверхности и их многомерные обобщения) размерности больше 1 понятие степени определяется не так, как это в одномерном случае сделано у нас, а аналогично тому выражению степени через число прообразов различных типов, которое в нашем случае будет указано в этом пункте.

Пусть $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ — непрерывное отображение. Допустим, что у некоторой точки $b \in \mathbb{S}^1$ имеется только конечное число прообразов $a_1, \ldots, a_n$. Мы распределим эти прообразы по трём различным типам и покажем, что степень отображения f определённым образом выражается через число прообразов различных типов. Точку $a_i \in f^{-1}(b)$ мы относим к одному из типов в зависимости от того, как расположены по отношению к b образы $f(z)$ близких к a_i чисел $z \in \mathbb{S}^1$ — а именно, от того, в какую сторону от b (против часовой стрелки или по ней) сдвигается $f(z)$, когда z слегка отклоняется от b в ту или иную сторону. Аккуратную формулировку удобно дать в терминах свойств угловой функции φ отображения f (какой-нибудь из многих его угловых функций).

Пусть α_i — один из аргументов комплексного числа a_i и $\beta_i = \varphi(\alpha_i)$ (это один из аргументов числа b). Поскольку прообразы числа b изолированы, то при достаточно малом $\varepsilon > 0$ из $x \in (\alpha_i - \varepsilon, \alpha_i + \varepsilon)$ и $x \neq \alpha_i$ следует, что $\varphi(x) \neq \beta_i$. Так как $\varphi(\alpha_i - \varepsilon, \alpha_i)$ и $\varphi(\alpha_i, \alpha_i + \varepsilon)$ суть связные множества, не содержащие β_i, то каждое из них целиком расположено по одну сторону от β_i (объясните это подробнее), т. е., например, либо все числа из $\varphi(\alpha_i - \varepsilon, \alpha_i)$ меньше β_i, либо все они больше β_i. Ради краткости условимся писать для подмножества $A \subset \mathbb{R}$, что $A < \beta_i$, соответственно, $A > \beta_i$, если все числа из A меньше β_i, соответственно, больше β_i. Будем говорить, что a_i является (для отображения f)

а) точкой подъёма, если $\varphi(\alpha_i - \varepsilon, \alpha_i) < \beta_i$ и $\varphi(\alpha_i, \alpha_i + \varepsilon) > \beta_i$,

б) точкой спуска, если $\varphi(\alpha_i - \varepsilon, \alpha_i) > \beta_i$ и $\varphi(\alpha_i, \alpha_i + \varepsilon) < \beta_i$,

в) точкой локального экстремума, если она не является ни точкой подъёма, ни точкой спуска, т. е. если $\varphi(\alpha_i - \varepsilon, \alpha_i)$ и $\varphi(\alpha_i, \alpha_i + \varepsilon)$ расположены по одну и ту же сторону от β_i.

Эта формулировка дана в терминах угловой функции $\varphi(x)$, но ясно, что при замене $\varphi(x)$ на другую угловую функцию $\varphi(x) + 2\pi k$, $k \in \mathbb{Z}$, точки подъёма останутся точками подъёма, точки спуска — точками спуска и точки локального экстремума — точками локального экстремума. (Множества $\varphi(\alpha_i - \varepsilon, \alpha_i)$ и $\varphi(\alpha_i, \alpha_i + \varepsilon)$ сдвинутся на $2\pi k$, и на столько же сдвинется β_i — теперь это будет $\varphi(\alpha_i) + 2\pi k$ = прежнее $\beta_i + 2\pi k$.) Равным образом ничего не изменится, если аргумент α_i комплексного числа a_i заменить на другой его аргумент $\alpha_i + 2\pi j$, $j \in \mathbb{Z}$. (Тогда β_j заменится на $\varphi(\alpha_i + 2\pi j) = \varphi(\alpha_i) + 2\pi lj$, где l — степень отображения f, а множества $\varphi(\alpha_i - \varepsilon, \alpha_i)$ и $\varphi(\alpha_i, \alpha_i + \varepsilon)$ тоже сдвинутся на $2\pi lj$.) Поэтому свойство точки a_i быть точкой подъёма, спуска или локального экстремума является свойством самой этой точки (говоря более педантично — свойством, относящимся к поведению отображения f возле этой точки) и никак не зависит от конкретного выбора привлекающихся при определении вспомогательных объектов — аргумента α_i комплексного числа a_i и угловой функции φ нашего отображения. Не зависит оно и от конкретного выбора «малого» положительного числа ε: ведь если $\varepsilon_1 > 0$ — другое такое число, что

$$\beta_i \notin \varphi(\alpha_i - \varepsilon_1, \alpha_i) \cap \varphi(\alpha_i, \alpha_i + \varepsilon_1)$$

и, скажем, $\varepsilon < \varepsilon_1$, то всё связное множество $\varphi(\alpha_i - \varepsilon_1, \alpha_i)$ находится по ту же сторону от β_i, что и его часть $\varphi(\alpha_i - \varepsilon, \alpha_i)$ (почему?), а $\varphi(\alpha_i, \alpha_i + \varepsilon_1)$ — по ту же сторону от β_i, что и $\varphi(\alpha_i, \alpha_i + \varepsilon)$.

Название «точка локального экстремума» согласуется со стандартным смыслом слов «локальный экстремум» (только при педантичном следовании стандартному словоупотреблению следовало бы сказать подробнее: «точка локального экстремума угловых функций»). Действительно, если a_i — точка локального экстремума в нашем смысле, то в зависимости от того, лежат ли оба множества $\varphi(\alpha_i - \varepsilon, \alpha_i)$ и $\varphi(\alpha_i, \alpha_i + \varepsilon)$ слева или, соответственно, справа от $\beta_i = \varphi(\alpha_i)$, для всех $x \in (\alpha_i - \varepsilon, \alpha_i + \varepsilon)$, отличных от α_i, будет выполняться одно и то же неравенство $\varphi(x) < \varphi(\alpha_i)$, соответственно, $\varphi(x) > \varphi(\alpha_i)$. Иными словами, α_i будет в обычном смысле слова точкой локального максимума функции φ, соответственно, точкой локального минимума этой функции, что и объединяется под общим названием «точка локального экстремума».

Названия же «точка подъёма» и «точка спуска» не являются стандартными. Они предлагаются для временного использования по аналогии со стандартными терминами «точка возрастания, соответственно, убывания», которые относятся к несколько менее общей ситуации. Именно, α_i называется точкой (локального) возрастания функции $\varphi(x)$, если у α_i имеется такая окрестность $(\alpha_i - \varepsilon, \alpha_i + \varepsilon)$, в которой функция φ возрастает, т. е. если $x, y \in (\alpha_i - \varepsilon, \alpha_i + \varepsilon)$ и $x < y$, то $\varphi(x) < \varphi(y)$. Ясно, что точка возрастания является точкой подъёма в нашем смысле, но обратное не обязательно, ибо в случае точки подъёма не исключено, что, скажем, справа от α_i и притом сколь угодно близко к α_i функция φ может колебаться — нужно только, чтобы при всех этих колебаниях она не становилась $\leqslant \varphi(\alpha_i)$. Аналогично, α_i называется точкой (локального) убывания функции $\varphi(x)$, если у α_i

имеется такая окрестность $(\alpha_i - \varepsilon,\ \alpha_i + \varepsilon)$, в которой функция φ убывает, т. е. если $x,\ y \in (\alpha_i - \varepsilon,\ \alpha_i + \varepsilon)$ и $x < y$, то $\varphi(x) > \varphi(y)$. Такая точка является точкой спуска в нашем смысле, но обратное не обязательно.

Выражение для степени отображения через число прообразов различных типов таково: *если имеется k точек подъёма и l точек спуска, то* $\mathrm{Ст}\, f = k - l$.

В доказательстве, естественно, показывается, что если φ — угловая функция для f, то $\varphi(x + 2\pi) - \varphi(x) = 2\pi(k - l)$. Достаточно установить это для какого-нибудь одного x (почему?). Таким x у нас будет аргумент одного из комплексных чисел $a_1,\ \ldots,\ a_n$. У каждого a_i имеется аргумент α_i, лежащий в полуинтервале $[0,\ 2\pi)$. Пронумеруем числа a_i таким образом, чтобы

$$0 \leqslant \alpha_1 < \alpha_2 < \ldots < \alpha_n < 2\pi.$$

Мы докажем, что $\varphi(\alpha_1 + 2\pi) - \varphi(\alpha_1) = 2\pi(k - l)$.

Присоединим к числам α_i ещё $\alpha_0 = \alpha_n - 2\pi,\ \ \alpha_{n+1} = \alpha_1 + 2\pi$, так что

$$\alpha_0 < 0 \leqslant \alpha_1 < \alpha_2 < \ldots < \alpha_n < 2\pi \leqslant \alpha_{n+1}, \tag{31}$$

и примем $\beta_i = \varphi(\alpha_i)$, $i = 0,\ \ldots,\ n+1$. На открытых интервалах $(\alpha_i,\ \alpha_{i+1})$ нет таких x, для которых $\varphi(x)$ было бы аргументом числа b (почему?), а значения β_i, β_{i+1} функции φ в концах такого интервала являются аргументами b. Поэтому $|\beta_{i+1} - \beta_i| \leqslant 2\pi$, — если бы β_{i+1} отличалось от β_i более чем на 2π, то в $(\alpha_i,\ \alpha_{i+1})$ имелась бы точка x, для которой $|\varphi(x) - \beta_i| = 2\pi$, т. е. для которой $\varphi(x)$ было бы аргументом b.

Теперь мы наведём некоторую формалистику, частично заменяя рассуждения формулами[45]; читателю полезно будет посмотреть на рисунках, изображающих несколько возможных графиков функции φ при небольших n, какие геометрические обстоятельства отражаются нашими формулами.

При $i = 1,\ \ldots,\ n$ обозначим

$$\varepsilon_i^- = \begin{cases} -1, & \text{если } \varphi(\alpha_i - \varepsilon,\ \alpha_i) < \beta_i, \\ \ \ 1, & \text{если } \varphi(\alpha_i - \varepsilon,\ \alpha_i) > \beta_i, \end{cases}$$

$$\varepsilon_i^+ = \begin{cases} -1, & \text{если } \varphi(\alpha_i,\ \alpha_i + \varepsilon) < \beta_i, \\ \ \ 1, & \text{если } \varphi(\alpha_i,\ \alpha_i + \varepsilon) > \beta_i. \end{cases}$$

[45] «Очерки по истории математики» Н. Бурбаки (это собрание исторических очерков из различных томов трактата «Элементы математики») заканчиваются словами, что великие мыслители-математики всегда стремились «вычисления заменять идеями». В русском переводе по ошибке написано обратное: «идеи заменять вычислениями». Смешно, что это тоже верно. Великому физику Н. Бору приписывают изречение, что глубокая истина — это такая, отрицание которой тоже является истиной.

В терминах $\varepsilon_i^{\pm}$ различные типы точек прообраза $f^{-1}(b)$ характеризуются так:

а) a_i — точка подъёма: $\varepsilon_i^- = -1$, $\varepsilon_i^+ = 1$;

б) a_i — точка спуска: $\varepsilon_i^- = 1$, $\varepsilon_i^+ = -1$;

в) a_i — точка локального экстремума: $\varepsilon_i^- = \varepsilon_i^+$.

Отсюда видно, что

$$\varepsilon_i^+ - \varepsilon_i^- = \begin{cases} 2 & \text{для точки подъёма,} \\ -2 & \text{для точки спуска,} \\ 0 & \text{для точки локального экстремума.} \end{cases} \tag{32}$$

Далее, если $\varepsilon_i^+ = -1$, то $\varphi(\alpha_i, \alpha_{i+1}) < \beta_i$ (соответствующее неравенство заведомо верно для близких к α_i точек из (α_i, α_{i+1}), а значит, и для всех точек из (α_i, α_{i+1}), поскольку там нет точек x с $\varphi(x) = \beta_i$). При этом для $\varphi(\alpha_{i+1}) = \lim\limits_{x\to\alpha_{i+1}} \varphi(x)$ будет либо $\varphi(\alpha_{i+1}) = \varphi(\alpha_i)$, либо $\varphi(\alpha_{i+1}) = \varphi(\alpha_i) - 2\pi$ (почему?). В первом случае при малых $\varepsilon > 0$ будет $\varphi(\alpha_{i+1} - \varepsilon, \alpha_{i+1}) < \beta_i = \varphi(\alpha_{i+1}) = \beta_{i+1}$, т. е. $\varepsilon_{i+1}^- = -1$; во втором случае $\varphi(\alpha_{i+1} - \varepsilon, \alpha_{i+1}) > \beta_{i+1}$, т. е. $\varepsilon_{i+1}^- = 1$. Заметим, что в первом случае $\frac{1}{2}(\varepsilon_i^+ - \varepsilon_{i+1}^-) = 0$, а во втором $\frac{1}{2}(\varepsilon_i^+ - \varepsilon_{i+1}^-) = -1$; поэтому в обоих случаях

$$\varphi(\alpha_{i+1}) = \varphi(\alpha_i) + \pi(\varepsilon_i^+ - \varepsilon_{i+1}^-). \tag{33}$$

Аналогично заключаем, что если $\varepsilon_i^+ = 1$, то $\varphi(\alpha_i, \alpha_{i+1}) > \beta_i$, и при этом либо $\beta_{i+1} = \varphi(\alpha_{i+1}) = \beta_i$, либо $\beta_{i+1} = \varphi(\alpha_{i+1}) = \beta_i + 2\pi$. В первом случае $\varepsilon_{i+1}^- = 1$, во втором $\varepsilon_{i+1}^- = -1$, и в обоих случаях выполняется (33). (Восполните детали.)

Осталось совсем недолго:

$$2\pi \operatorname{Ст} f = \varphi(\alpha_1 + 2\pi) - \varphi(\alpha_1) = \varphi(\alpha_{n+1}) - \varphi(\alpha_1) = \sum_{i=1}^{n} (\varphi(\alpha_{i+1}) - \varphi(\alpha_i)) =$$

$$= \sum_{i=1}^{n} \pi(\varepsilon_i^+ - \varepsilon_{i+1}^-) = \pi \sum_{i=1}^{n} \varepsilon_i^+ - \pi \sum_{i=1}^{n} \varepsilon_{i+1}^- = \pi \sum_{i=1}^{n} \varepsilon_i^+ - \pi \sum_{i=2}^{n+1} \varepsilon_i^-.$$

В последней сумме ε_{n+1}^- можно заменить на ε_1^-, потому что φ одинаково ведёт себя возле точек a_{n+1} и a_1 (поясните это подробнее). Поэтому

$$2\pi \operatorname{Ст} f = \pi \sum_{i=1}^{n} \varepsilon_i^+ - \pi \sum_{i=1}^{n} \varepsilon_i^- = \pi \sum_{i=1}^{n} (\varepsilon_i^+ - \varepsilon_i^-).$$

Согласно (32), те i, для которых a_i — точки локального экстремума, не дают вклада в сумму. Вклад каждой из точек подъёма равен 2π, а вклад каждой точки спуска равен -2π. Стало быть, сумма равна умноженной на 2π разности $k-l$ между числом точек подъёма и числом точек спуска, что и требовалось доказать.

Замечание. Определяя три разных типа точек прообраза, мы пользовались одной и той же координатой φ и возле рассматриваемой точки прообраза a_i, и возле её образа b. Но может оказаться удобным использовать какие-то другие координаты ξ и η, определённые, возможно, не на всей окружности, а только в некоторой окрестности точки a_i (координата ξ) и в некоторой окрестности точки b (координата η; в таких случаях говорят о локальных координатах). В координате ξ некоторая открытая дуга C, содержащая a_i, изображается интервалом (ξ', ξ''), причём a_i имеет координату ξ_i. В координате η некоторая открытая дуга D, содержащая b, изображается интервалом (η', η''), причём b имеет координату η_0. Соответствие

$$C \leftrightarrow (\xi', \xi''), \qquad D \leftrightarrow (\eta', \eta''),$$
$$\text{точка дуги} \leftrightarrow \text{координата } \xi \text{ или } \eta$$

является гомеоморфизмом. Непрерывное отображение $f|C$ (т. е. f, рассматриваемое только на дуге C) описывается некоторым непрерывным отображением $\psi\colon (\xi', \xi'') \to (\eta', \eta'')$. (Так как координаты локальные, ни о какой многозначности ψ или об изменении ψ при приращении аргумента, соответствующем обороту по окружности, не возникает речи — делая оборот по $\mathbb{S}^1$, мы выйдем из области определения локальной координаты.)

При использовании локальных координат мы по-прежнему можем различать три случая, соответствующие точкам подъёма, спуска и локального экстремума. Но может случиться, что направление возрастания локальной координаты противоположно положительной ориентации $\mathbb{S}^1$ (направлению возрастания φ). Если, скажем, для ξ это направление противоположно, а для η — нет, то точка, для которой $\psi(\xi_i-\varepsilon, \xi_i) < \eta_0$, $\psi(\xi_i, \xi_i+\varepsilon) > \eta_0$, будет уже не точкой подъёма, а точкой спуска. Ведь дуга окружности с координатами из $(\xi_i-\varepsilon, \xi_i)$ примыкает к точке α_i в направлении возрастания φ и потому интервал $(\xi_i-\varepsilon, \xi_i)$ в терминах φ соответствует некоторому интервалу вида $(\alpha_i, \alpha_i+\varepsilon')$; и так далее. Аналогично, точка, для которой

$$\psi(\xi_i-\varepsilon, \xi_i) > \eta_0, \quad \psi(\xi_i, \xi_i+\varepsilon) < \eta_0,$$

будет точкой подъёма. Точки же локального экстремума по-прежнему характеризуются тем, что $\psi(\xi_i-\varepsilon, \xi_)$ и $\psi(\xi_i, \xi_i+\varepsilon)$ лежат по одну сторону от η_0. (Другое дело, что, скажем, точка локального минимума в терминах ξ, η выглядит как точка локального максимума.) Аналогичные изменения свойственны и тому случаю, когда направление возрастания ξ соответствует положительному направлению на $\mathbb{S}^1$, а направление возрастания η — нет.

Наконец, если оба эти направления (возрастания ξ и η) соответствуют отрицательной ориентации $\mathbb{S}^1$, характеризация трёх типов прообразов не меняется (почему?).

§ 5. Векторные поля

Похвальное слово векторным полям

Если каждой точке некоторого $A \subset \mathbb{R}^n$ сопоставлен некоторый вектор (который при этом обычно представляют себе как направленный отрезок с началом в этой точке), то говорят, что на A задано векторное поле (см. рис. 33), а если сопоставлено некоторое число, то говорят, что задано скалярное поле (когда векторы как бы противопоставляют числам, то последние называют скалярами). Если A — область, то говорят о поле в области. Как видно, понятие скалярного поля, по существу, совпадает с понятием функции, уточним: функции с числовыми значениями. А векторное поле — это, по существу, функция, только принимающая векторные значения.

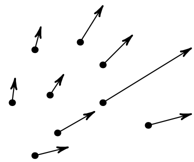


Рис. 33. Векторное поле

Если математическое понятие поля равносильно понятию функции, то зачем употребляют два названия для одного и того же объекта? Это делают в связи с тем, как данный объект возник и (или) что в связи с этим мы собираемся с ним делать, так что данный вопрос — не формального определения, а смыслового оттенка.

Причин, побуждающих говорить именно о полях, несколько. Одна из них состоит в том, что математические поля[46)] часто возникают при изучении физических полей. С последними нам не придётся иметь дело, но я всё-таки скажу о них несколько слов. Из повседневной жизни нам хорошо известно, что такое вещество. Физика XVIII уточнила различные свойства вещества, а потом (окончательно уже в XX в.) было доказано, что привычное нам «сплошное» вещество состоит из атомов, которые объединяются в молекулы. Но оказывается, что имеются не менее реальные физические объекты — поля, основными примерами которых являются поле тяготения и электромагнитное поле. (В микромире имеются и другие поля, но от нашей повседневной жизни это далеко.) Привычное проявление поля состоит в том, что возле некоего вещественного объекта («источника поля») на другие кусочки вещества действуют некоторые силы. (В случае электромагнитного поля вещество, о котором идёт речь, должно нести электрические заряды. Иногда они взаимно компенсируются, но при

46) Не только векторные или скалярные. Бывает, что точкам сопоставляются не векторы и не скаляры, а объекты несколько иной природы. Мы не будем иметь с ними дела.

попадании в электромагнитное поле всё-таки проявляются.) Сперва это пытались описывать как прямое действие одних объектов на другие. Такое описание для сил гравитации даётся законом Ньютона. Для электрического поля имеется аналогичный закон Кулона, но этим дело не кончается: на движущиеся заряды действуют ещё магнитные силы. Попытки выразить закон их действия в каком-то аналогичном виде не увенчались успехом. Зато увенчалась полным успехом идея, что в пространстве, окружающем источник поля, происходят какие-то изменения, там имеется нечто, воздействующее на помещённые в эту область кусочки вещества. Мы воспринимаем воздействие как действие источника поля на эти кусочки, а на самом деле непосредственно на них действует созданное таким источником поле. (Поле имеется и там, где никаких кусочков вещества нет, хотя мы этого не замечаем.)

Рассмотрим поле тяготения, создаваемое некоторым телом или некоторыми телами. Напряжённость $\boldsymbol{E} = \boldsymbol{E}(A)$ этого поля в точке A — это сила, с которой данное поле действует на помещённую в точку A «пробную» единичную массу. Таким образом, физическому объекту — полю тяготения — сопоставляется векторное поле напряжённости. Оказывается, что, зная последнее, можно вычислить силу, действующую в данном поле на любую массу, даже не обязательно сосредоточенную в точке. (Аналогичная ситуация имеет место и для электрического поля. Надо, однако, сказать, что на самом деле электрическое поле связано с магнитным, образуя единое электромагнитное поле, и соответствующий математический образ сложнее, чем векторное поле.)

Впрочем, наиболее наглядный пример векторного поля не совсем таков — он связан с веществом. Именно, если мы имеем дело с течением жидкости или газа в некоторой области, то каждой её точке соответствует вектор скорости той частицы жидкости, которая в данный момент времени находится в этой точке. Полученное векторное поле называется вектором скоростей (данного течения); оно может изменяться со временем. А простейший пример скалярного поля — это поле температур. Имеется некое нагретое тело; каждой его точке сопоставляется число — температура в этой точке; это и есть интересующее нас поле. В некоторых случаях изучение векторных полей сводится к изучению вспомогательных скалярных полей, уже не столь наглядных, но математически более простых. Я не буду этого объяснять, но если читатель достаточно знаком с электростатикой, он может подумать о поле потенциала.

Математическое описание физического поля отчасти напоминает математическое описание вещества в случае сплошной среды, и поэтому долгое время пытались объяснить электромагнитное поле как проявление процессов, происходящих в некоей особой всепроникающей среде, «эфире». Но оказалось, что этому эфиру приходится приписывать свойства, качественно отличные от свойств всех веществ, и его никак нельзя обнаружить. Пришлось признать поле «первичной реальностью» наравне с веществом (тем более, что, как выяснилось, многие привычные свойства вещества являются не какими-то «первичными» свойствами, а следствиями электромагнитного взаимодействия составляющих это вещество атомов — т. е. полевых процессов! — в сочетании с особенностями квантовой механики — механики микрообъектов).

Если векторное поле $\boldsymbol{u}(A)$ служит для математического описания физического поля, то физически неестественно «отрывать» вектор $\boldsymbol{u}(A)$ от точки A. Его лучше так и представлять себе — исходящим из A. В математике тоже бывает, что лучше представлять себе вектор $\boldsymbol{u}(A)$ «исходящим» из A. Например, рассмотрим на единичной окружности $\mathbb{S}^1$ векторное поле, в котором точке z сопоставляется то же самое z, рассматриваемое уже как вектор. Формально речь идёт о функции $z \mapsto z$, но нагляднее рисовать вектор, идущий из z в $2z$. Ясно, что это векторное поле можно непрерывно продолжить с окружности на её внешнюю область таким образом, чтобы оно нигде не обращалось в нуль — простейший вариант даётся той же формулой $z \mapsto z$, и пока что для функции с векторными (или, как мы условились поступать в случае плоскости, с комплексными) значениями это выглядит не сложнее, чем для векторного поля. А вот при непрерывном продолжении во внутреннюю область поле обязательно будет где-то обращаться в нуль. Об этом можно догадаться, поэкспериментировав с рисунками, а с чем бы мы стали экспериментировать, говоря о функции? В доказательстве же удобно попеременно говорить то о поле, то о функции.

Ещё одна причина, побуждающая говорить о полях, состоит в том, что имеется несколько понятий, специфических именно для поля. Для полей физического происхождения эти понятия имеют физический смысл и обычно фигурируют в тех или иных физических законах. Одно из них, возможно, памятно читателю из школьного курса физики: поток векторного поля через поверхность. Другое — циркуляция векторного поля вдоль некоторой замкнутой линии, — возможно, ему не известно, но для физики оно не менее важно. С ними тесно связаны некоторые «дифференциальные операции» (дивергенция и ротор), которые в терминах координат определённым образом выражаются через производные координат рассматриваемого вектора. Те же понятия имеют смысл и для «математических» полей, будучи в этом случае, конечно, уже просто некоторыми математическими объектами. Формально ничто не мешает рассматривать эти объекты для какой-нибудь векторной функции, заданной в некоторой области. Но тогда непонятно, почему именно им надо уделять особое внимание. Если обращение к такому объекту действительно оказывается целесообразным по характеру рассматриваемого вопроса, то часто бывает уместно называть исходную векторную функцию полем. Это сразу вызывает в памяти комплекс представлений и результатов, традиционно относимых к математической теории векторных полей, что тоже может оказаться полезным.

Мы не будем иметь дела с этими понятиями, но у нас встретятся другие понятия, которые тоже выглядят естественными и довольно наглядными для векторных полей, а для векторных функций не кажутся таковыми. Они имеют иной характер и возникли в самой математике, а не в физике (однако со временем оказались не совсем бесполезными для последней, см. приложение к [6]).

Топологические понятия, относящиеся к векторным полям

Пусть Σ^1 — ориентированная топологическая окружность, на которой задано непрерывное векторное поле $\boldsymbol{v}$, нигде не обращающееся в нуль. Вращением $\mathrm{Вр}_{\Sigma^1}\,\boldsymbol{v}$ этого поля на Σ^1 называется степень отображения

$$\Sigma^1 \to \mathbb{S}^1, \qquad z \mapsto \frac{\boldsymbol{v}(z)}{|\boldsymbol{v}(z)|}.$$

Собственно, в этом определении не важно, что у Σ^1 нет самопересечений — всё равно в конечном счёте речь идёт о степени отображения $\mathbb{S}^1 \xrightarrow{g} \Sigma^1 \xrightarrow{v/|v|} \mathbb{S}^1$, где g — «параметризация» замкнутой кривой, которая выше была топологической окружностью, но которая вместо этого вполне могла бы иметь самопересечения. Только при их наличии выбор параметризации становится более существенным (от него зависит уже не только знак степени). Если упорно держаться по возможности ближе к традиционным геометрическим понятиям, то пришлось бы обратиться к уже упоминавшемуся подходу, изложенному в [8]. Но в духе топологии проще поступить иначе (что даёт и бо́льшую свободу в обращении с вращением векторного поля).

(Непрерывная) петля в $\mathbb{R}^n$ или в множестве $M \subset \mathbb{R}^n$ — это отображение $f\colon \mathbb{S}^1 \to \mathbb{R}^n$ или $\mathbb{S}^1 \to M$. Практически совершенно эквивалентное понятие — замкнутый путь, т. е. такой путь $g\colon I \to M$, у которого начало и конец совпадают: $g(0) = g(1)$. Эквивалентность состоит в том, что если f — петля, то $g(t) = f(e^{2\pi i t})$ — замкнутый путь, а если g — замкнутый путь, то $f(z) = g\left(\frac{1}{2\pi}\arg z\right)$ — петля. Выражение «вращение непрерывного поля $\boldsymbol{v}$, не обращающегося в нуль ни в одной точке замкнутой кривой, на этой кривой» получает удачное уточнение как вращение $\mathrm{Вр}_f\,\boldsymbol{v}$ на петле (вдоль петли) $f\colon \mathbb{S}^1 \to \mathbb{C}$, по определению являющееся степенью отображения

$$\mathbb{S}^1 \xrightarrow{f} \text{область определения поля } \boldsymbol{v} \neq 0 \xrightarrow{v/|v|} \mathbb{S}^1$$

(см. рис. 34). Это соответствует сделанному выше общему замечанию, что мы часто работаем с путями и допускаем достаточно широкий класс их деформаций, поэтому нам незачем обращать внимание на различные тонкости с кривыми (необходимые в других вопросах). Если мы говорим

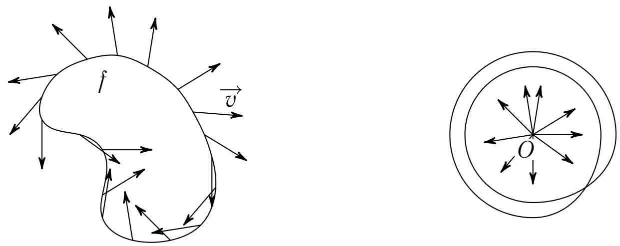


Рис. 34. На петле f задано векторное поле $\boldsymbol{v}$. Перенеся начала всех векторов в 0 и сделав их векторами единичной длины (с прежними направлениями), видим, что $\mathrm{Bp}_f\,\boldsymbol{v} = 2$.

о петлях, то гомотопия (или деформация) петли — это гомотопия именно непрерывного отображения $f\colon \mathbb{S}^1 \to M$. А в терминах замкнутых путей это такая гомотопия

$$F\colon I \times I \to M, \qquad F(s,\,0) = f(s) \text{ при всех } s \in I,$$

при которой путь всё время остаётся замкнутым, т. е. $F(0,\,t) = F(1,\,t)$ (в других обозначениях, $f_t(0) = f_t(1)$) при всех t. Из определения ясно, что $\mathrm{Bp}_f\,\boldsymbol{v}$ не меняется при гомотопии петли f в той области, где определено и отлично от нуля поле $\boldsymbol{v}$. (С соответствующим отображением $\mathbb{S}^1 \to \mathbb{S}^1$ тоже происходит гомотопия.) Интуитивно понятно, что при любом разумном определении эквивалентности двух параметризаций замкнутой кривой, «определяющих одно и то же направление на ней», соответствующие петли будут гомотопными[47]; поэтому гомотопическая точка зрения (как уже говорилось) является более общей и гибкой, будучи одновременно и более простой.

Пусть f — петля на плоскости $\mathbb{C}$, и пусть во всех точках этой петли (т. е. во всех точках $f(\mathbb{S}^1)$) задано непрерывное векторное поле $\boldsymbol{v}$, нигде не обращающееся в нуль, причём $\mathrm{Bp}_f\,\boldsymbol{v} \neq 0$. Оказывается, что такое поле невозможно продолжить до нигде не обращающегося в нуль непрерывного векторного поля на всей плоскости $\mathbb{C}$. Действительно, допустим, что такое

[47] Во всяком случае, это так, если понимать эквивалентность параметризаций в духе [8]. Однако доказывать данное утверждение нам незачем, ибо для наших целей определение кривой из [8] не привлекается.

продолжение $\boldsymbol{w}$ существует. Рассмотрим отображения

$$g_t\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad g_t(z) = \frac{\boldsymbol{w}(tf(z))}{|\boldsymbol{w}(tf(z))|} \qquad (t \in I,\ z \in \mathbb{S}^1).$$

Это гомотопия, в которой g_1 — то самое отображение, которое фигурирует в определении $\mathrm{Вр}_f\, \boldsymbol{v}$, а g_0 — отображение в одну точку $\dfrac{\boldsymbol{w}(0)}{|\boldsymbol{w}(0)|}$. Следовательно, $\mathrm{Ст}\, g_t = 0$ и, в частности, $\mathrm{Вр}_f\, \boldsymbol{v} = \mathrm{Ст}\, g_1 = 0$.

Если петля — это «параметризация» топологической окружности Σ^1, то можно доказать больше: *продолжение $\boldsymbol{v}$ с Σ^1 до некоторого непрерывного и нигде не равного нулю поля $\boldsymbol{w}$ во внутреннюю область возможно в том и только том случае, когда* $\mathrm{Вр}_{\Sigma^1}\, \boldsymbol{v} = 0$, *продолжение же во внешнюю область всегда возможно*. Сама формулировка подразумевает известной теорему Жордана.

Упражнение 25. Докажите выделенное курсивом утверждение в частном случае, когда $\Sigma^1 = \mathbb{S}^1_r(A)$ является окружностью радиуса r с центром в какой-нибудь точке A. (Указание. При продолжении во внешнюю область можно принять $\boldsymbol{w}$ постоянным на лучах, начинающихся в A. Если $\mathrm{Вр}_{\mathbb{S}^1_r(A)}\, \boldsymbol{v} = 0$, то обозначим через $\{g_t\}$ гомотопию, соединяющую фигурирующее в определении $\mathrm{Вр}_{\mathbb{S}^1_r(A)}\, \boldsymbol{v}$ отображение

$$g_1\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad g_1(z) = \frac{\boldsymbol{v}(A + rz)}{|\boldsymbol{v}(A + rz)|}$$

с отображением в точку, и положим

$$\boldsymbol{w}(w) = f_{|w-A|/r}\Bigl(\frac{w - A}{|w - A|}\Bigr).)$$

Утверждение, что при определённых условиях векторное поле $\boldsymbol{w}(A)$ с координатами $(f(x, y), g(x, y))$ (здесь (x, y) — координаты точки A) где-то обязательно обращается в нуль, обеспечивает существование решения у системы уравнений

$$f(x, y) = 0, \quad g(x, y) = 0.$$

Подробнее об этой стороне дела говорится в [4].

Порядок $\mathrm{Пор}(f, A)$ петли $f\colon \mathbb{S}^1 \to \mathbb{C}$ относительно точки $A \notin f(\mathbb{S}^1)$, или порядок точки A относительно петли f — это степень отображения

$$\mathbb{S}^1 \to \mathbb{S}^1, \qquad z \mapsto \frac{f(z) - A}{|f(z) - A|}.$$

Очевидно, $\mathrm{Пор}(f, A)$ уточняет смысл выражения «число оборотов, которые делает вектор $\overrightarrow{AB}$, когда точка $B = f(z)$ делает один обход по петле». $\mathrm{Пор}(f, A)$ не меняется при такой деформации петли f, при которой она никогда не проходит через A (т. е., в понятных обозначениях, $A \notin F(\mathbb{S}^1 \times I)$).

Пусть $f\colon \mathbb{S}^1 \to \mathbb{C}$ — петля, не проходящая через 0, и $\mathrm{Пор}(f, 0) \neq 0$. Повернём эту петлю на угол α вокруг 0 в положительном направлении, т. е. возьмём петлю $f_1 = e^{i\alpha} f$. Докажем, что эти две петли (т. е. их образы) обязательно пересекаются: $f(\mathbb{S}^1) \cap f_1(\mathbb{S}^1) \neq \varnothing$.

Допустим, они не пересекаются. Пользуясь полярными координатами, напишем

$$f(e^{i\varphi}) = r(\varphi)e^{i\psi(\varphi)}, \quad f_1(e^{i\varphi}) = r(\varphi)e^{i\psi(\varphi)+i\alpha},$$

где $r(\varphi)$ — периодическая функция с периодом 2π, а угловая функция $\psi(\varphi)$ обладает свойством $\psi(\varphi + 2\pi) = \psi(\varphi) + 2\pi k$, $k \neq 0$.

Пусть $r(\varphi)$ достигает минимума при каком-то $\varphi = \varphi_{\min}$ и максимума — при $\varphi = \varphi_{\max}$. Отрезок $J = \{tf_1(e^{i\varphi_{\min}});\ 0 \leqslant t \leqslant 1\}$ соединяет точку $f_1(e^{i\varphi_{\min}})$ с 0. При $0 \leqslant t < 1$ модуль $|tf_1(e^{i\varphi_{\min}})| < \min\limits_{\varphi} r(\varphi) = \min\limits_{\varphi} |f(e^{i\varphi})|$, так что при этих t точка $tf_1(e^{i\varphi_{\min}}) \notin f(\mathbb{S}^1)$. А когда $t = 1$, то $f_1(e^{i\varphi_{\min}}) \notin f(\mathbb{S}^1)$ по предположению. Таким образом, отрезок J не пересекается с $f(\mathbb{S}^1)$. Поэтому $\text{Пор}(f, f_1(e^{i\varphi_{\min}})) = \text{Пор}(f, 0) = k \neq 0$. Луч $\{tf_1(e^{i\varphi_{\max}});\ t \geqslant 0\}$ начинается в точке $f_1(e^{i\varphi_{\max}})$, которая по предположению $\notin f(\mathbb{S}^1)$, а все прочие точки луча отстоят от 0 дальше, чем любая точка $f(\mathbb{S}^1)$. Поэтому луч не пересекается с $f(\mathbb{S}^1)$. Но ясно, что при больших t, когда из точки $tf_1(e^{i\varphi_{\max}})$ весь образ $f(\mathbb{S}^1)$ виден под небольшим углом, $\text{Пор}(f, tf_1(e^{i\varphi_{\max}})) = 0$. Значит, и $\text{Пор}(f, f_1(e^{i\varphi_{\max}})) = 0$. Но при непрерывном изменении φ от $\varphi_{\min}$ до $\varphi_{\max}$ точка $f_1(e^{i\varphi})$ никогда не попадает на $f(\mathbb{S}^1)$; следовательно, $\text{Пор}(f, f_1(e^{i\varphi}))$ при этом должен оставаться постоянным, тогда как мы видели, что при $\varphi = \varphi_{\min}$ он равен $k \neq 0$, а при $\varphi = \varphi_{\max}$ он равен 0.

Индекс особой точки векторного поля

Содержание настоящего пункта не используется в дальнейшем. Но оно непосредственно примыкает к предыдущему и само по себе имеет важное значение, поэтому уместно сделать соответствующее отступление от основной линии изложения.

Пусть в области G задано непрерывное векторное поле $\boldsymbol{v}$. С топологической точки зрения, особого внимания заслуживают нули этого поля (точки, где $\boldsymbol{v} = 0$). Ведь если $\boldsymbol{v}(A) = 0$, то $\boldsymbol{v}$ не определяет никакого направления в точке A, между тем как в любой другой точке $\boldsymbol{v}$ определяет некоторое направление (сквозь «направленный отрезок» проходит некоторая прямая). Поэтому о нулях поля $\boldsymbol{v}$ часто говорят как о его особых точках, хотя возможно, что с аналитической точки зрения никаких особенностей там нет, т. е. координаты вектора $\boldsymbol{v}(B)$ являются «хорошими» функциями координат точки B — непрерывными, гладкими, аналитическими, даже многочленами. С другой стороны, «особая точка» A может быть действительно особой в том смысле, что вектор $\boldsymbol{v}$ в ней не определён, а при приближении B к A вектор $\boldsymbol{v}(B)$ не стремится к пределу (который можно было бы принять за $\boldsymbol{v}(A)$).

Ниже A — изолированная особая точка поля $\boldsymbol{v}$, заданного и непрерывного в некоторой области G (возможно, за исключением этой точки). Подразумевается, стало быть, что вектор $\boldsymbol{v}$ определён во всех близких к A точках $B \neq A$, — скажем, при $\rho(B, A) < \varepsilon_0$, — и что в них он задаёт некоторое направление, т. е. $\boldsymbol{v}(B) \neq 0$. Рассмотрим вращение $\mathrm{Вр}_{\mathbb{S}^1_\varepsilon(A)}\, \boldsymbol{v}$ этого поля на окружности $\mathbb{S}^1_\varepsilon(A)$ (это по-прежнему окружность радиуса ε с центром в A). При всех $\varepsilon \in (0, \varepsilon_0)$ оно одно и то же: если $\varepsilon_1, \varepsilon_2 \in (0, \varepsilon_0)$, то $\mathrm{Вр}_{\mathbb{S}^1_{\varepsilon_1}(A)}\, \boldsymbol{v}$ и $\mathrm{Вр}_{\mathbb{S}^1_{\varepsilon_2}(A)}\, \boldsymbol{v}$ суть степени отображений $\mathbb{S}^1 \to \mathbb{S}^1$, соединяемых гомотопией

$$(z, t) \mapsto \frac{\boldsymbol{v}(A + \varepsilon_1 + t(\varepsilon_2 - \varepsilon_1))z)}{|\boldsymbol{v}(A + \varepsilon_1 + t(\varepsilon_2 - \varepsilon_1))z)|}.$$

Поэтому число $\mathrm{Вр}_{\mathbb{S}^1_\varepsilon(A)}\, \boldsymbol{v}$ определяется самой особой точкой A (т. е., конечно, не точкой как таковой, а поведением поля $\boldsymbol{v}$ в сколь угодно малой её окрестности). Оно называется индексом $\mathrm{Инд}\, A$ особой точки A; если надо специально указать, что речь идёт о поле $\boldsymbol{v}$, то можно говорить об индексе особой точки A поля $\boldsymbol{v}$ и писать $\mathrm{Инд}(A, \boldsymbol{v})$.

В определении индекса вместо окружностей можно использовать кривые гораздо более общего вида. Пусть $f\colon \mathbb{S}^1 \to \mathbb{C}$ — петля, которая (т. е. образ которой) не содержит точки A и целиком содержится в круге $\{w;\ |w-A| < \varepsilon_0\}$ (с упомянутым выше ε_0; всюду в этом круге поле определено и непрерывно, причём $\boldsymbol{v} = 0$ только в точке A), и пусть $\mathrm{Пор}(f, A) = 1$. Тогда $\mathrm{Инд}(A, \boldsymbol{v}) = \mathrm{Вр}_f\, \boldsymbol{v}$. Действительно, возьмём какое-нибудь $\varepsilon \in (0, \varepsilon_0)$ и рассмотрим гомотопию

$$f_t\colon \mathbb{S}^1 \to \mathbb{C}, \quad f_t(z) = A + (1-t)(f(z) - A) + t\varepsilon\frac{f(z) - A}{|f(z) - A|} \quad (t \in I,\ \varepsilon \in \mathbb{S}^1),$$

соединяющую петлю $f = f_0$ с петлёй $f_1\colon z \mapsto A + \varepsilon\dfrac{f(z) - A}{|f(z) - A|}$ и не проходящую через A. Последняя же петля гомотопна петле $z \mapsto A + \varepsilon z$ посредством гомотопии, тоже не проходящей через A. Ведь, согласно определению $\mathrm{Пор}(f, A)$, отображение

$$g_0\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad g_0(z) = \frac{f(z) - A}{|f(z) - A|}$$

имеет степень 1, а потому гомотопно тождественному отображению $1_{\mathbb{S}^1}$ (оставляющему все точки $\mathbb{S}^1$ на месте). Пусть $\{g_t\}$ — соответствующая гомотопия (g_0 имеет прежний смысл, $g_1 = 1_{\mathbb{S}^1}$). В наших обозначениях петля f_1 есть $z \mapsto A + \varepsilon g_0(z)$; гомотопия $h_t(z) = A + \varepsilon g_t(z)$ связывает её с петлёй $h_1\colon z \mapsto A + \varepsilon z$ и не проходит через A. Эти две гомотопии связывают f с последней петлёй и не проходят через A, поэтому $\mathrm{Пор}(f, A) = \mathrm{Пор}(h_1, A)$. Но вращение поля $\boldsymbol{v}$ на петле h_1 есть $\mathrm{Вр}_{\mathbb{S}^1_\varepsilon(A)}\, \boldsymbol{v} = \mathrm{Инд}(A, \boldsymbol{v})$.

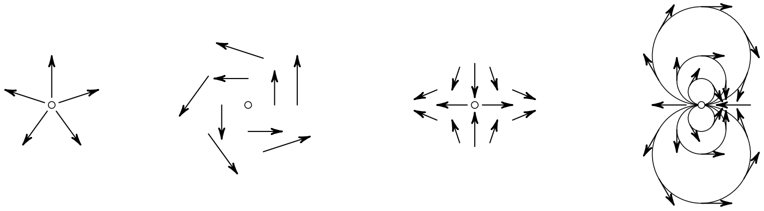


Рис. 35

На рис. 35 изображено несколько различных особых точек и указаны их индексы.

Упражнение 26. Формально определение индекса дословно годится для любой неособой точки поля (заданного и непрерывного в некоторой области). Докажите, что у неособой точки индекс равен нулю, так что и говорить о нём незачем.

Сформулируем без доказательства следующую теорему.

Теорема. *Пусть топологическая окружность* Σ^1 *ограничивает внутреннюю область* W, *и пусть в* $\operatorname{clos} W = W \cup \Sigma^1$ *задано непрерывное векторное поле* $\boldsymbol{v}$, *имеющее только конечное число особых точек, причём ни одна из них не находится на* Σ^1. *Тогда* $\text{Вр}_{\Sigma^1} \boldsymbol{v} = \sum\limits_A \text{Инд}(A, \boldsymbol{v})$.

Возникает вопрос, нельзя ли, зная только $\text{Вр}_{\Sigma^1} \boldsymbol{v}$, сказать ещё что-нибудь об особых точках поля $\boldsymbol{v}$? Оказывается, нет:

Пусть на Σ^1 *задано непрерывное векторное поле* $\boldsymbol{v}$, *нигде не обращающееся в нуль. Пусть* $A_1, \ldots, A_l$ — *любые точки в* W *и* $k_1, \ldots, k_l$ — *любые целые числа, причём* $\sum k_i = \text{Вр}_{\Sigma^1} \boldsymbol{v}$. *Тогда существует такое непрерывное продолжение* $\boldsymbol{w}$ *поля* $\boldsymbol{v}$ *в* $\operatorname{clos} W$, *что* $\boldsymbol{w} = 0$ *только в точках* A_i *и* $\text{Инд}(A_i, \boldsymbol{w}) = k_i$.

§ 6. Некоторые применения к алгебре

Основная теорема алгебры

В качестве первого применения материала §§ 4—5 докажем так называемую **основную теорему алгебры**: *любой многочлен* $f(z)$ *ненулевой степени имеет хотя бы один корень в* $\mathbb{C}$.

Из этой теоремы в алгебре выводится, что $f(z)$ имеет столько корней, какова его степень, с учётом кратности этих корней; это означает, например, что если степень $f(z)$ равна n, то $f(z)$ представляется в виде $a_n(z - z_1) \ldots (z - z_n)$, где a_n — старший коэффициент многочлена $f(z)$ (коэффициент при z^n), а $z_1, \ldots, z_n$ — корни многочлена, причём в этом списке одно и то же число может повторяться несколько раз, а именно, каждый корень фигурирует в этом списке столько раз, какова его кратность. Но соответствующее рассуждение уже не имеет отношения к нашей теме и потому опускается; я только замечу, что оно на самом деле главное содержится в утверждении об одном корне, а остальное значительно проще.

Основная теорема алгебры была сформулирована как предположение голландским математиком А. Жираром в начале XVII века, в дальнейшем в течение почти двух столетий уверенность в её справедливости возрастала, но полностью доказать её смог только К. Ф. Гаусс в 1799 г. Такой большой промежуток времени между возникновением гипотезы и её доказательством объясняется, конечно, не недостатком способностей или интереса у предшественников Гаусса, а тем, что необходимые для доказательства средства не так уж просты и на их, так сказать, «вызревание» ушло немало времени. Имеется несколько доказательств основной теоремы алгебры, основанных на различных идеях и использующих различные «технические» средства. (Уже сам Гаусс предложил несколько доказательств.) В этой книге, естественно, приводится доказательство, использующее излагаемый в ней аппарат — степень отображения, вращение векторного поля. Оформление этого аппарата относится к ещё более позднему времени — оно произошло на грани XIX и XX вв., но фактически уже в начале XIX в. в работах О. Коши сформировалась как самостоятельное научное направление теория функций комплексного переменного, в рамках которой можно провести аналогичное доказательство.

Название «основная теорема алгебры» является исторической условностью. Даже в 1799 г. в алгебре было много важных вещей, никак не зависящих от «основной теоремы», а после того развития, какое эта дисциплина получила за прошедшие с тех пор два века, за «основной» теоремой остаётся роль предложения довольно начального характера, важного для значительной части, но всё же только части алгебры. Пожалуй, более интересно особое положение «основной теоремы алгебры» среди прочих сведений, относящихся к подобного рода начальному материалу: там это, по существу, единственное утверждение, доказательство которого принципиально требует не-алгебраических средств, заимствованных или из математического анализа (в широком смысле слова, включая теорию функций комплексного переменного), или, как у нас, из геометрии и топологии. (На более высоком уровне привлечение средств, выходящих за пределы собственно алгебры, является, естественно, обычным делом.)

Пусть $f(z) = a_n z^n + a_{n-1} z^{n-1} + \ldots + a_1 z + a_0$, где $a_n \neq 0$. Разделив $f(z)$ на a_n, можно ограничиться многочленами с единичным старшим коэффициентом, т. е. многочленами вида

$$f(z) = z^n + a_{n-1}z^{n-1} + \ldots + a_1 z + a_0.$$

Будем рассматривать $f(z)$ как векторное поле (комплексные числа обычным образом рассматриваются как векторы и точке $z \in \mathbb{C}$ сопоставляется вектор $f(z)$). (Ни полужирным шрифтом, ни стрелкой мы при этом не пользуемся.) Достаточно установить, что вращение $\mathrm{Bp}_{\mathbb{S}^1_r} f$ этого поля на окружности достаточно большого радиуса r с центром в нуле отлично от 0; тогда внутри окружности поле f должно хотя бы в одной точке обращаться в нуль. Доказательство основано на том, что при достаточно большом r на окружности $|z| = r$ в выражении для $f(z)$ слагаемое z^n больше по модулю всех остальных слагаемых, вместе взятых. В основном именно это слагаемое оно определяет направление вектора $f(z)$. Поэтому вращение $\mathrm{Bp}_{\mathbb{S}^1_r} f$ оказывается таким же, каким оно было бы для $f_0(z) = z^n$. А $\mathrm{Bp}_{\mathbb{S}^1_r} f_0 = n \neq 0$.

Приступая к реализации этой идеи, начнём с середины — с того места, где говорится, что вращение $\mathrm{Bp}_{\mathbb{S}^1_r} f$ определяется «преобладающим» первым слагаемым z^n. В более общем виде речь идёт о следующем утверждении, в формулировке которого буква f имеет иной смысл, а векторные поля снова обозначаются полужирным шрифтом, так как оно не связано специфически с многочленами. Пусть на петле $f\colon \mathbb{S}^1 \to \mathbb{C}$ (т. е. на её образе $f(\mathbb{S}^1)$) задано ненулевое (нигде не обращающееся в нуль) векторное поле $\boldsymbol{v}$ и векторное поле $\boldsymbol{w}$, причём всюду $|\boldsymbol{w}| < |\boldsymbol{v}|$. Тогда вращения векторных полей $\boldsymbol{v}$ и $\boldsymbol{v} + \boldsymbol{w}$ на петле f совпадают:

$$\mathrm{Bp}_f \boldsymbol{v} = \mathrm{Bp}_f(\boldsymbol{v} + \boldsymbol{w}).$$

Для доказательства рассмотрим семейство векторных полей $\boldsymbol{v}_t = \boldsymbol{v} + t\boldsymbol{w}$, $t \in I$ (по-прежнему $I = [0, 1]$). В духе § 4 семейство $\{\boldsymbol{v}_t;\ t \in I\}$ можно назвать гомотопией векторных полей (хотя применительно к векторным полям слово «гомотопия» почему-то не очень употребительно), имея в виду, что $\boldsymbol{v}_t(z)$ непрерывно зависит от (z, t). Такая гомотопия связывает поля $\boldsymbol{v}_0 = \boldsymbol{v}$ и $\boldsymbol{v}_1 = \boldsymbol{v} + \boldsymbol{w}$. Поскольку $|\boldsymbol{v}(z)| > |\boldsymbol{w}(z)|$ при всех $z \in f(\mathbb{S}^1)$, то нигде на петле $f(\mathbb{S}^1)$ ни одно из полей $\boldsymbol{v}_t$ не обращается в нуль, и мы можем говорить об их вращениях на этой петле $\mathrm{Bp}_f \boldsymbol{v}_t$. Это суть степени отображений

$$g_t\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad z \mapsto \frac{\boldsymbol{v}_t(f(z))}{|\boldsymbol{v}_t(f(z))|},$$

при которых $g_t(z)$ непрерывно зависит от (z, t), так что $\{g_t\}$ является гомотопией, соединяющей g_0 и g_1. Значит, при всех t степени этих отображений одни и те же. А при $t = 0$ и $t = 1$ степени равны $\mathrm{Bp}_f \boldsymbol{v}$ и $\mathrm{Bp}_f(\boldsymbol{v} + \boldsymbol{w})$, соответственно.

Применим данное соображение к вращениям фигурирующих выше полей $f_0(z)$ и $f(z)$ на окружности $\mathbb{S}^1_r$ большого радиуса r. Так как $f(z) = f_0(z) + g(z)$, где $g(z) = a_{n-1}z^{n-1} + \ldots + a_1 z + a_0$, то нам нужны такие r, чтобы при $|z| = r$ было $|z^n| > |g(z)|$. Ясно, что это обеспечено, если $r > 1 + |a_{n-1}| + \ldots + |a_1| + |a_0|$ — ведь тогда

$$1 < r^{n-1}, \quad |z| < r^{n-1}, \quad \ldots, \quad |z|^{n-1} = r^{n-1}$$

(пока использовано, что $r > 1$) и

$$\begin{aligned}|g(z)| \leqslant |a_{n-1}z^{n-1}| + \ldots + |a_1 z| + |a_0| \leqslant \\ \leqslant |a_{n-1}|r^{n-1} + \ldots + |a_1|r^{n-1} + |a_0|r^{n-1} = \\ = (|a_{n-1}| + \ldots + |a_1| + |a_0|)r^{n-1} < r^n.\end{aligned}$$

Остаётся напомнить, что угловая функция отображения

$$\mathbb{S}^1 \to \mathbb{S}^1_r \xrightarrow{f_0} \mathbb{C} \setminus \{0\} \to S^1$$
$$z = e^{i\varphi} \mapsto w = rz \mapsto w^n \mapsto \frac{w^n}{|w^n|} = \frac{r^n e^{in\varphi}}{r^n} = e^{in\varphi}$$

есть $\varphi \mapsto n\varphi$ и потому его степень, т. е. $\mathrm{Bp}_{\mathbb{S}^1_r} f_0$, равна n.

Вещественные корни многочленов

Займёмся теперь вещественными корнями многочленов. Естественно, многочлены при этом предполагаются имеющими вещественные коэффициенты (такие многочлены короче называют вещественными многочленами). Ниже это часто подразумевается само собой разумеющимся и не оговаривается особо.

На вопрос о наличии вещественных корней не может быть столь же простого общего ответа, какой в комплексной области даётся основной теоремой алгебры — одни вещественные многочлены имеют вещественные корни, другие (например, $x^2 + 1$) — нет. Но можно указать правило, позволяющее определить, имеет ли данный многочлен вещественные корни или нет, и если имеет, то сколько.

Для формулировки этого правила нам понадобятся так называемые цепные (или непрерывные) дроби. Что это такое, проще всего пояснить на примере, относящимся не к многочленам, а к натуральным числам.

Рассмотрим дробь $\frac{23}{17}$. Имеем

$$\frac{23}{17} = 1 + \frac{6}{17} = 1 + \frac{1}{17/6}, \qquad \frac{17}{6} = 2 + \frac{5}{6} = 2 + \frac{1}{6/5}, \qquad \frac{6}{5} = 1 + \frac{1}{5},$$

поэтому (идя назад)

$$\frac{17}{6}=2+\cfrac{1}{1+\cfrac{1}{5}},\qquad \frac{23}{17}=1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{5}}}.$$

Подобные «многоэтажные» выражения и называются цепными дробями. Давать (довольно громоздкое) формальное определение этому понятию, надеюсь, излишне.

Делить друг на друга с остатком можно не только натуральные числа, но и многочлены. Например:

$$\begin{array}{r|l}
x^5+6x^3\phantom{{}-x^2}+x+1 & x^3+1 \\ \cline{2-2}
-\;x^5\phantom{{}+6x^3}+x^2\phantom{{}+x+1} & x^2+6 \\ \cline{1-1}
6x^3-x^2+x+1 & \\
-\;6x^3\phantom{{}-x^2+x}+6 & \\ \cline{1-1}
-x^2+x-5 &
\end{array}\;.$$

Поэтому

$$\frac{x^5+6x^3+x+1}{x^3+1}=(x^2+6)+\frac{-x^2+x-5}{x^3+1}=(x^2+6)+\frac{1}{(x^3+1)/(-x^2+x-5)}.$$

На следующем шаге мы осуществляем в знаменателе последнего выражения деление с остатком:

$$\begin{array}{r|l}
x^3\phantom{{}-x^2+5x}+1 & -x^2+x-5 \\ \cline{2-2}
-\;x^3-x^2+5x\phantom{{}+1} & -x-1 \\ \cline{1-1}
x^2-5x+1 & \\
-\;x^2-x+5 & \\ \cline{1-1}
-4x-4 &
\end{array}$$

и потому

$$\frac{x^3+1}{-x^2+x-5}=(-x-1)+\frac{-4x-4}{-x^2+x-5}=(-x-1)+\frac{1}{(-x^2+x-5)/(-4x-4)}.$$

Далее,

$$\begin{array}{r|l}
-x^2+x-5 & -4x-4 \\ \cline{2-2}
-\;-x^2-x\phantom{{}-5} & \frac{1}{4}x-\frac{1}{2} \\ \cline{1-1}
2x-5 & \\
-\;2x+2 & \\ \cline{1-1}
-7 &
\end{array}\;,$$

$$\frac{-x^2+x-5}{-4x-4} = \frac{1}{4}x - \frac{1}{2} + \frac{-7}{-4x-4} = \\ = \frac{1}{4}x - \frac{1}{2} + \frac{1}{(-4x-4)/(-7)} = \frac{1}{4}x - \frac{1}{2} + \frac{1}{\frac{4}{7}x + \frac{4}{7}},$$

после чего мы можем написать

$$\frac{x^3+1}{-x^2+x-5} = (-x-1) + \cfrac{1}{\left(\frac{1}{4}x - \frac{1}{2}\right) + \cfrac{1}{\frac{4}{7}x + \frac{4}{7}}},$$

и окончательно

$$\frac{x^5+6x^3+x+1}{x^3+1} = (x^2+6) + \cfrac{1}{(-x-1) + \cfrac{1}{\left(\frac{1}{4}x + \frac{1}{2}\right) + \cfrac{1}{\frac{4}{7}x + \frac{4}{7}}}}.$$

Как известно, отношение двух многочленов (с не равным тождественно нулю знаменателем) называется рациональной функцией. Соображения такого же характера, как для конкретной рациональной функции $\dfrac{x^5+6x^3+x+1}{x^3+1}$ выше, приводят к выводу, что любую рациональную функцию можно единственным образом представить в виде цепной дроби

$$f_0 + \cfrac{1}{f_1 + \cfrac{1}{f_2 + \cfrac{1}{f_3 + \ddots \; f_{k-1} + \cfrac{1}{f_k}}}}, \tag{34}$$

где степени многочленов $f_1, \ldots, f_k$ положительны, а многочлен f_0 может иметь и нулевую степень (т. е. сводиться к константе, даже быть нулём[48]). Если исходная рациональная функция вещественна, т. е. если она является отношением двух вещественных многочленов, то все f_i тоже вещественны.

[48] Многочлен f_0 равен 0, если f является таким отношением двух многочленов $\frac{g}{h}$, что степень g меньше степени h — в этом случае мы сразу «переворачиваем» дробь и начинаем делить h на g. f_0 является ненулевой константой, если степени g и h равны. При вычислении любого другого f_i мы делим некоторый многочлен на многочлен меньшей степени, поэтому степени этих f_i положительны.

Обозначим для вещественного многочлена $g = a_n x^n + a_{n-1}x^{n-1} + \ldots$ $\ldots + a_1 x + a_0$ степени n (тем самым сказано, что $a_n \neq 0$)

$$j(g) = \begin{cases} 0, & \text{если } n \text{ чётное},\\ 1, & \text{если } n \text{ нечётное и } a_n > 0,\\ -1, & \text{если } n \text{ нечётное и } a_n < 0. \end{cases}$$

Теорема. *Пусть f — вещественный многочлен. Представим рациональную функцию* $\frac{f'}{f}$ (где f' — производная многочлена f) *в виде цепной дроби*

$$\frac{f'}{f} = \cfrac{1}{f_1 + \cfrac{1}{f_2 + \cfrac{1}{f_3 + \ddots + \cfrac{1}{f_{k-1} + \cfrac{1}{f_k}}}}} \tag{35}$$

(в отличие от (34), здесь отсутствует слагаемое f_0, ибо степень f' меньше степени f). *Тогда число вещественных корней многочлена f равно*

$$j(f_1) - j(f_2) + \ldots + (-1)^{k-1} j(f_k).$$

(*Здесь речь идёт о числе корней без учёта их кратности, так что корень кратности* 10 *засчитывается всего один раз, как и однократный корень.*)

Эту теорему и приводимое ниже её доказательство мне сообщил А. Г. Хованский. Надо сказать, что ещё в 1829 г. французский математик Ж. Ш. Ф. Штурм указал другое правило для определения числа вещественных корней вещественного многочлена f. Имеются модификации обоих правил, относящиеся к числу корней, лежащих в заданном отрезке (собственно, теорему Штурма обычно сразу так и формулируют. Помимо учебников высшей алгебры, она изложена в популярной брошюре [9] [49].), но в этой книжке я ограничиваюсь только простейшим случаем, когда этот отрезок совпадает с $\mathbb{R}$.

Правила Хованского и Штурма эквивалентны — в них даются различные способы нахождения одного и того же числа. Естественно, что каждое

[49] Теорема эта справедливо считалась большим достижением и обеспечила Штурму известность среди математиков. Рассказывают, что на своих лекциях Штурм юмористически говорил о ней: «теорема, имя которой я имею честь носить».

Часто (в том числе и в [9]) при изложении теоремы Штурма с самого начала предполагают, что многочлен f не имеет кратных корней. Такое предположение не ограничивает общности, ибо, как объясняют в алгебре, от кратных корней легко избавиться, но всё же стоит заметить, что на самом деле теорема Штурма верна и при наличии кратных корней.

из этих правил можно вывести из другого, причём, в соответствии с чисто алгебраическим характером формулировок этих правил, вывод можно провести уже чисто алгебраическим путём. Применительно к задаче о числе всех вещественных корней, которой мы ограничиваемся в настоящей книжке, это несложно. Но поскольку я вообще не останавливаюсь на правиле Штурма, я не могу подробно обсуждать его связи с правилом Хованского.

Замечу, что теорему Штурма можно доказать при помощи соображений топологического характера, хотя и не таких, как у Хованского, но по существу родственных. (Обычно же её доказывают иначе). Я выбрал для изложения теорему Хованского, потому что её доказательство всё-таки проще. Связывая её с именем Хованского, я исхожу из того, что эта теорема и метод её доказательства, по-видимому, являются новыми. Правда, литература по расположению корней многочленов обширна, и нельзя поручиться, что мы с Хованским знаем всё существенное, что там есть. С другой стороны, теорема и доказательство Хованского являются побочным результатом или частным случаем проведённого им исследования более сложных современных вопросов, и нет особых сомнений в новизне его подхода. А поскольку возникновение последнего было связано с современными вопросами, то вполне возможно, что в более простой задаче о подобном подходе не возникало мысли (хотя в такой задаче он совсем прост).

В доказательстве правила Хованского используется геометрическое понятие, вероятно, известное части читателей — вещественная проективная прямая. Она является простейшим понятием так называемой проективной геометрии и, как и вся эта геометрия, само по себе понятие проективной прямой возникло и стало играть заметную роль в науке в первой половине XIX в. совершенно независимо от топологии (которой тогда и не было) или вещественных корней многочленов. Если освоиться с этим понятием, то остальное совсем просто. Но боюсь, что даже среди студентов младших курсов, которые отчасти знакомы с проективной геометрией, не у всех имеется достаточная ясность в этом вопросе. Поэтому я счёл нужным специально остановиться на том, что такое вещественная проективная прямая.

С чисто топологической точки зрения так называемая проективная прямая — это самая обыкновенная окружность. Но часто бывает, что один и тот же объект может играть различные роли. Когда мы его рассматриваем вместе с этими ролями, получаются уже как бы различные объекты. А если в связи с этими ролями объекты снабжаются некоторыми дополнительными структурами, то комбинации «объект плюс дополнительная структура» — это уже новые (как бы «обогащённые») объекты, и из-за различия дополнительных структур они могут быть различными уже без всяких «как бы».

Формально нам из всей этой дополнительной структуры, которая и делает топологическую окружность проективной прямой, нужно очень немногое. Пусть ψ — обычная угловая координата на $\mathbb{S}^1$ (как всегда, отсчитываемая по модулю 2π). Тогда в качестве координат на некоторых частях $\mathbb{S}^1$ мы используем также $u = -\operatorname{tg}(\psi/2)$ и $v = -\operatorname{ctg}(\psi/2)$. (Знак минус взят потому, что он получается, когда u и v вводятся посредством описываемых ниже геометрических построений.) Координата u определена там, где $\psi \not\equiv \pi \bmod 2\pi$ (т. е. $\psi \neq k\pi$ с нечётным k), а v — там, где $\psi \not\equiv 0 \bmod 2\pi$ (т. е. $\psi \neq 2k\pi$ с целым k). Эти координаты будут использоваться при определении и исследовании некоторых вспомогательных отображений проективной прямой на себя.

Но такой формальный подход оставляет в стороне мотивировку того, что делается, хотя, в конечном счёте, надо было бы мотивировать только введение этих u, v и определение упомянутых отображений. Очень естественные конструкции при этом выглядят как трюки. Чтобы у читателя не сложилось подобного превратного впечатления, я останавливаюсь подробнее на том, что же такое проективная прямая на самом деле. Мой рассказ об этом не совсем удовлетворителен с педагогической точки зрения, потому что проективная прямая — это одно из нескольких тесно связанных друг с другом понятий проективной геометрии, которая «по-настоящему» начинается не с проективной прямой, а с проективной плоскости. Но это увело бы нас слишком далеко, а то, что говорится ниже о проективной прямой, всё-таки лучше, чем ничего.

Проективная прямая

Точки стандартной окружности $\mathbb{S}^1$ суть концы единичных векторов, приложенных к началу координат O; если угодно, $\mathbb{S}^1$ образована самими этими единичными векторами (мы ссылаемся на обычное отождествление точек с векторами, когда имеется фиксированное «начало отсчёта»). Любой такой вектор определяет направление на плоскости $\mathbb{R}^2$. С другой стороны, направление можно охарактеризовать ещё так: во-первых, задаётся некоторая прямая на плоскости (заранее не требуется, чтобы прямая проходила через O, но если она не проходит через O, можно взять параллельную ей прямую, проходящую через O); во-вторых, на этой прямой указывается некоторое «положительное» направление или, как ещё говорят, ориентация [50]. Таким образом, направления, о которых мы говорим в связи с $\mathbb{S}^1$ — это направления ориентированных прямых на плоскости. При желании можно было бы считать, что $\mathbb{S}^1$ — это совокупность всех таких направлений, или ещё совокупность всех прямолинейных лучей, на-

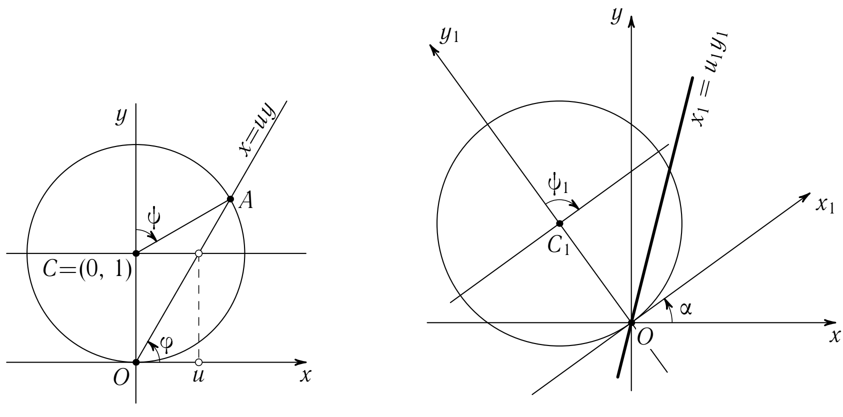
[50] Ранее говорилось об ориентациях на окружности и указывалось, что по существу там тоже речь идёт в выборе направления, принимаемого за положительное.

чинающихся в O. В настоящей книжке $\mathbb{S}^1$ часто играет именно такую роль, а окружность единичного радиуса с центром в O привлекается скорее как удобный способ представить себе эту совокупность в виде некоторой геометрической фигуры. Но если бы мы всерьёз приняли такой подход за исходный, то надо было бы объяснить, почему и как мы можем применять к этой совокупности топологические понятия, т. е. какие множества направлений (лучей) считаются открытыми подмножествами $\mathbb{S}^1$ (а после этого надо было бы выяснить, в какой степени при этом $\mathbb{S}^1$ наделяется «хорошими» топологическими свойствами). Если же $\mathbb{S}^1$ с самого начала появляется как соответствующая фигура на плоскости, то подобных вопросов не возникает.

Мы говорили о направлениях ориентированных прямых. Но ведь можно говорить и просто о направлениях прямых, не ориентируя последних. По определению, совокупность всех таких направлений — это и есть проективная прямая $\mathbb{RP}^1$. (С какой стати её называют «прямой», будет объяснено позднее.) В обозначении $\mathbb{RP}^1$ буква $\mathbb{P}$ указывает, что это не обычная прямая $\mathbb{R}$, а проективная; верхний индекс указывает на размерность (у $\mathbb{RP}^1$ имеются n-мерные «родственники» — так называемые n-мерные проективные пространства $\mathbb{RP}^n$, с которыми мы не будем иметь дела); а $\mathbb{R}$ указывает, что речь идёт о вещественной проективной прямой (подобно тому как наряду с $\mathbb{R}^n$ имеются комплексные пространства $\mathbb{C}^n$, — мы их не рассматриваем, но читатель вполне мог о них слышать, — так в проективной геометрии наряду с $\mathbb{RP}^n$ имеются комплексные проективные пространства $\mathbb{CP}^n$, с которыми мы тем более не будем иметь дела). Геометрический образ, непосредственно отвечающий сказанному — это совокупность всевозможных прямых на плоскости, проходящих через начало координат O. (Аналогичным геометрическим образом для совокупности направлений ориентированных прямых служит совокупность всевозможных лучей, начинающихся в O.) Он довольно нагляден; непривычно только то, что сейчас мы рассматриваем каждую такую прямую как некий единый объект — «точку проективной прямой». Несоответствие с Евклидом (согласно которому «точка есть то, что не имеет частей») ещё можно довольно легко переварить; хуже, что пока $\mathbb{RP}^1$ — это только множество некоторых элементов (прямых), не снабжённое никакой структурой.

Впрочем, хотя в формальном определении множества $\mathbb{RP}^1$ об этом не говорится, наш геометрический образ кое-что подсказывает насчёт структуры. По крайней мере, наглядно совершенно ясен смысл утверждения «прямые l_n стремятся при $n \to \infty$ к прямой l». (Угол между l и l_n должен стремится к нулю.) Раз так, то кажется правдоподобным, что в $\mathbb{RP}^1$ можно ввести топологические понятия, начиная с исходного для них понятия открытого подмножества. Это действительно возможно и совсем не сложно. Одна беда — это требует места. Поэтому мы пойдём другим путём.

Экономный способ действия состоит в том, чтобы подобрать геометрическую фигуру F, которая давала бы $\mathbb{RP}^1$ естественное геометрическое представление в том же смысле, как окружность единичного радиуса с центром в O даёт геометрическое представление совокупности начинающихся в O лучей (тем самым и совокупности направлений ориентированных прямых) на плоскости. Точки фигуры F должны «естественным образом» соответствовать прямым, проходящим через O; «естественность» этого соответствия призвана сделать интуитивно убедительным перенесение с его помощью топологических понятий, заведомо осмысленных применительно к F (ведь F — подмножество $\mathbb{R}^2$), в $\mathbb{RP}^1$, где у нас таких понятий в чётко оформленном виде пока нет (хотя предыдущий геометрический образ навевает какие-то смутные представления такого рода).



а) б)

Рис. 36

В качестве такой фигуры F годится окружность $\mathbb{S}^1_1(0, 1)$ единичного радиуса с центром в точке $C = (0, 1)$ (см. рис. 36, а)). Заметим, что эта окружность проходит через O. Прямая l, проходящая через O и отличная от оси абсцисс, пересекает эту окружность в самой точке O и ещё в одной точке A. Соответствие $l \leftrightarrow A$ — это и есть нужное нам «естественное» соответствие между элементами $\mathbb{RP}^1$ и точками настоящей геометрической фигуры $\mathbb{S}^1_1(C)$. Пока что оно не совсем взаимно однозначно — ось x была исключена из рассмотрения, а точка O окружности $\mathbb{S}^1_1(C)$ не соответствует никакой прямой. Но когда прямая l стремится к горизонтальной оси, точка A стремится к O, поэтому естественно принять, что точка O соответствует оси абсцисс; тем самым мы доопределяем соответствие между прямыми и точками $\mathbb{S}^1_1(C)$ до уже взаимно однозначного соответствия $\mathbb{RP}^1 \leftrightarrow \mathbb{S}^1_1(C)$.

Упражнение 27. а) Точки z окружности $\mathbb{S}^1$ имеют угловые координаты φ — полярные углы этих точек или, на другом языке, аргументы z как комплексных чисел. На окружности $\mathbb{S}^1_1(C)$ естественно использовать аналогичную угловую координату ψ, приняв за полюс центр C этой окружности, а за начальный луч — луч, начинающийся в C и лежащий на положительной полуоси ординат. Проверьте, что прямая, пересекающая $\mathbb{S}^1$ в точке с угловой координатой φ, пересекает $\mathbb{S}^1_1(C)$ в точке с угловой координатой $\psi = 2\varphi - \pi$. (Та же прямая пересекает $\mathbb{S}^1$ также в диаметрально противоположной точке с угловой координатой $\varphi+\pi$. Для ψ получается $2\varphi + \pi$. Это другая угловая координата той же самой точки окружности $\mathbb{S}^1_1(C)$, как оно, конечно, и должно быть.)

б) Наш выбор $\mathbb{S}^1_1(C)$ в качестве фигуры F, изображающей $\mathbb{RP}^1$, содержит несколько больше произвола, чем совсем уж самоочевидный выбор окружности $\mathbb{S}^1$ в качестве фигуры, изображающей множество лучей с началом в O. Ведь мы могли бы с равным основанием взять вместо $\mathbb{S}^1_1(C)$ любую другую окружность единичного радиуса, проходящую через O, т. е. окружность $\mathbb{S}^1_1(C_1)$ с центром в точке $C_1 = (\cos\alpha, \sin\alpha)$ с любым (вещественным) α (см. рис. 36, б)). В таком случае в качестве угловой координаты на $\mathbb{S}^1_1(C_1)$ естественно использовать полярный угол ψ_1, получающийся, когда центр C_1 этой окружности принимается за полюс, а начальным лучом служит луч, лежащий на полупрямой OC_1, начинающийся в C_1 и являющийся продолжением отрезка OC_1. Докажите, что $\psi_1 \equiv \psi + \pi - 2\alpha \bmod 2\pi$. ($\equiv$ и $\mathrm{mod}\ 2\pi$ означают, что это — сравнение по модулю 2π, т. е. разность его левой и правой частей кратна 2π. При переходе в факторгруппу $\mathbb{R}/\mathbb{Z}$ получается уже настоящее равенство.) Отсюда следует (как?), что топологические свойства $\mathbb{RP}^1$ не изменятся, если мы перенесём их в $\mathbb{RP}^1$ не из $\mathbb{S}^1_1(C)$, а из $\mathbb{S}^1_1(C_1)$.

Пора объяснить, какое отношение проективная прямая $\mathbb{RP}^1$ имеет к обычной прямой $\mathbb{R}$. Почти любая прямая l, проходящая через O, кроме оси абсцисс, имеет уравнение $x = uy$ с постоянным «угловым коэффициентом» u, своим для каждой прямой. (Обычно уравнение прямой, проходящей через O, пишут в виде $y = kx$ с постоянным k, которое и называется угловым коэффициентом. У нас x и y поменялись ролями.) В частности, прямая l пересекает прямую $R = \{(x, y); y = 1\}$ в точке $(u, 1)$. Можно представить себе, что прямая R получается из оси абсцисс (как обычно, отождествляемой с $\mathbb{R}$) при подъёме последней вверх на 1. Так или иначе, получается соответствие

$$\mathbb{RP}^1 \setminus \{\text{ось } x\} \leftrightarrow \mathbb{R}, \qquad l \leftrightarrow u. \tag{36}$$

Ось абсцисс параллельна прямой R и, конечно, не может её пересекать. Но когда $u \to \pm\infty$, прямые $x = uy$ стремятся к оси абсцисс. Поэтому можно доопределить соответствие (36) до взаимно однозначного, присоединив к $\mathbb{R}$ символ ∞ и сопоставив его оси абсцисс. Итак, мы имеем соответствие

$$\mathbb{R} \cup \{\infty\} \leftrightarrow \mathbb{RP}^1 \leftrightarrow \mathbb{S}^1_1(0, 1),$$

которое определяется с помощью описанных геометрических конструкций с добавлением, что

$$\infty \leftrightarrow \{\text{ось } x\} \leftrightarrow \text{ точка } O \in \mathbb{S}^1_1(0,\,1) \quad (\text{для неё } \psi \equiv \pi \bmod 2\pi).$$

Упражнение 28. Как связаны u и ψ?

Таким образом, проективная прямая получается из обычной прямой при добавлении одного-единственного нового элемента ∞, который, как подсказывает обозначение, называется «бесконечно удалённой» (или ещё «несобственной») точкой. Обратите внимание, что добавлена только одна несобственная точка. С точки зрения элементарного курса анализа или неэлементарной теории функций вещественного переменного, было бы естественнее добавить два несобственных элемента, $\infty = +\infty$ и $-\infty$. При этом из $\mathbb{R}$ получается множество, которое в понятном смысле можно обозначить через $[-\infty,\,\infty]$ и которое (при разумном введении топологических понятий) оказывается гомеоморфным обычным отрезкам (замкнутым интервалам) вроде $[0,\,1]$. Но в проективной геометрии к $\mathbb{R}$ добавляется только один несобственная точка, и при этом, как мы знаем, получается топологическая окружность. Это пример того, что в различных случаях бывает целесообразно по-разному добавлять несобственные элементы — пример далеко не единственный, но, вероятно, один из самых простых.

Исторически вначале проективная прямая $\mathbb{RP}^1$ как раз и определялась путём дополнения точки ∞ к обычной прямой $\mathbb{R}$. (Точнее, несобственные элементы добавлялись к обычным плоскости и пространству, что давало $\mathbb{RP}^2$ и $\mathbb{RP}^3$. Специально прямой тогда не занимались, но было ясно, что к каждой прямой на плоскости при переходе от $\mathbb{R}^2$ к $\mathbb{RP}^2$ автоматически добавляется ровно одна бесконечно удалённая точка.) У такого образа действий имеются два недостатка. Один нам по существу уже известен, хотя и по другому поводу: надо всё определять заново. Другой состоит в том, что поначалу точка ∞ играет какую-то особую роль. Хотя в этой книжке об основном содержании проективной геометрии нет речи — она затрагивается только постольку, поскольку нам нужна проективная прямая, что означает привлечение ничтожной и не особенно показательной части этой дисциплины, — я всё же скажу, что одним из основных мотивов при возникновении последней было стремление не только ввести бесконечно удалённые элементы, но и обеспечить возможность работать с ними (в соответствующих задачах) как с обычными точками (нам это тоже понадобится в пределах наших скромных потребностей), а такого «равноправия» обычных и несобственных точек нельзя достичь быстро, если поначалу несобственные элементы занимают какое-то особое положение. Между тем, при нашем исходном определении $\mathbb{RP}^1$ как совокупности направлений прямых на плоскости или, более формально, как совокупности прямых, проходящих через O, все элементы $\mathbb{RP}^1$ выглядят совершенно равноправными — чем горизонтальное направление или ось x отличается от других направлений или от проходящих через O прямых? Особое положение оси x было введено искусственно, когда мы взяли специальным образом расположенную окружность $\mathbb{S}^1_1(0,\ 1)$ и особенно когда взяли прямую R, параллельную этой оси.

С другой точки зрения, сопоставление (36) чисел u элементам l проективной прямой, отличным от оси x, — это некоторая координата в области $\mathbb{RP}^1 \setminus \{\text{ось } x\}$ этой прямой. Эта область получается из $\mathbb{RP}^1$ удалением одной-единственной точки (оси x). Удаляя какую-нибудь другую точку (прямую l_1), получим другую область, в которой с помощью аналогичного построения можно ввести координату u_1. А именно, будем рассматривать l_1 как ось абсцисс новой декартовой координатной системы, получающейся при повороте исходных координатных осей на некоторый угол α, и повторим в новой системе те построения, которые делались раньше.

Упражнение 29. а) Покажите, что координаты u и u_1 получаются друг из друга с помощью некоторого дробно-линейного преобразования, т. е.

$$u_1 = \frac{au + b}{cu + d} \tag{37}$$

с некоторыми постоянными a, b, c, d, причём определитель

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \neq 0$$

(что позволяет, обратно, выразить u через u_1). Как связано это преобразование с углом α?

б) Помимо поворота координатных осей, имеется ещё очевидное координатное преобразование — отражение, при котором начало координат остаётся на месте, на одной из осей координат положительное направление заменяется на противоположное, а на другой оси оно не меняется. Проделав такое преобразование, повторим затем предыдущее построение, вводящее координату в некоторой области на проективной прямой. Что это будет за область и как связана новая координата с прежней?

в) Наконец, можно взять произвольную замену декартовых координат, при которой начало координат остаётся на месте. (При предыдущих заменах оси координат оставались перпендикулярными друг другу и масштабные векторы на осях не менялись. Такие преобразования называются ортогональными, хотя это название не отражает сохранения масштабов.) Затем, как обычно, вводим локальную координату в некоторой области проективной прямой. Как связана эта координата с прежней координатой u?

г) Можно ли таким путём получить в некоторой области на $\mathbb{RP}^1$ замену координат вида (37) с наперёд заданными a, b, c, d, удовлетворяющими условию $ad - bc \neq 0$?

д) Какой специальный вид имеет преобразование координат описанного выше типа $u \to u_1$, если области определения u и u_1 совпадают, т. е. получаются при удалении из $\mathbb{RP}^1$ одной и той же точки?

Координаты, описанные в упражнении 29 (в частности, введённая выше координата u), играют особую роль в проективной геометрии — примерно такую же, какую обычные декартовы координаты играют в рассуждениях, касающихся обычной плоскости $\mathbb{R}^2$. Они называются аффинными координатами на проективной прямой (подразумевается — не на всей $\mathbb{RP}^1$, а на $\mathbb{RP}^1$ с одной выброшенной точкой). Название вызвано следующим. Аффинная геометрия на обычной прямой, плоскости или в пространстве — это раздел геометрии, посвящённый тем свойствам геометрических фигур на прямой, плоскости или в пространстве, которые не меняются при так называемых аффинных преобразованиях. Последние — это такие преобразования, которые в терминах декартовых координат выражаются как неоднородные линейные преобразования[51)]. Аффинные координаты на прямой, плоскости или в пространстве — это обычные декартовы координаты (с произвольными началом координат, осями, которые могут образовывать друг с другом углы, отличные от 90°, и произвольными масштабами по осям). Говоря об аффинных координатах в $\mathbb{RP}^1 \setminus \{\text{одна точка}\}$, мы тем самым указываем, что по удалении одной точки (условно принимаемой за «бесконечно удалённую») оставшаяся часть $\mathbb{RP}^1$ становится обычной прямой с обычной аффинной структурой, в понятном смысле соответствующей этим координатам. Теперь можно завершить описание проективной структуры на $\mathbb{RP}^1$, отличающей $\mathbb{RP}^1$ от гомеоморфных ей топологических окружностей: один из (эквивалентных друг другу) вариантов этого понятия состоит в том, что множество $\mathbb{RP}^1$ рассматривается вместе с аффинными координатами, каждая из которых определена в области, получающейся удалением из $\mathbb{RP}^1$ одной точки, и которые выражаются друг через друга при помощи дробно-линейных преобразований.

Из всего этого нам в дальнейшем понадобится не так уж много. Нам понадобится введённая выше координата u в области $\mathbb{RP}^1 \setminus \{\text{ось } x\}$ и другая координата v, которая аналогичным образом вводится в области $\mathbb{RP}^1 \setminus \{\text{ось } y\}$. Напомню, что первая координата для прямой l является коэффициентом в уравнении этой прямой $x = uy$. Вторая координата для той же прямой — это коэффициент в уравнении этой прямой $y = vx$. Отсюда ясно, что $v = \dfrac{1}{u}$. В терминах «$\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$» вторая координата определена на множестве $(\mathbb{R} \setminus \{0\}) \cup \{\infty\}$. На $(\mathbb{R} \setminus \{0\})$ она равна $\dfrac{1}{u}$, а в точке ∞ она равна нулю. В терминах второй координаты окрестности точки ∞ — это множества, содержащие открытые интервалы $(-\varepsilon, \varepsilon)$. В терминах u такой открытый интервал есть $\left\{u;\ |u| > \dfrac{1}{\varepsilon}\right\} \cup \{\infty\}$.

[51)] Для плоскости и пространства имеется простая и красивая геометрическая характеризация аффинных преобразований в чисто геометрических терминах, без обращения к координатам: это такие взаимно однозначные преобразования плоскости (пространства) на себя, при которых прямые линии переходят в прямые линии. (Но доказать эквивалентность такого определения данному выше далеко не просто.) Для прямой такая характеризация, конечно, не годится.

Любая рациональная функция f определяет непрерывное отображение [52]

$$f\colon \mathbb{R} \setminus \{\text{полюса } f\} \to \mathbb{R}.$$

Оказывается, что, если рассматривать $\mathbb{R}$ как $\mathbb{RP}^1 \setminus \{\infty\}$, то это отображение единственным образом доопределяется до непрерывного отображения $\mathbb{RP}^1 \to \mathbb{RP}^1$, которое мы по-прежнему будем обозначать через f.

Действительно, пусть $f = \dfrac{g}{h}$, где многочлены g и h не имеют общих корней. В терминах проективной прямой мы имеем отображение

$$\mathbb{RP}^1 \setminus (\{\text{нули } h\} \cup \{\infty\}) \to \mathbb{RP}^1$$

(точка с координатой u переходит в точку с u-координатой $f(u)$). Пусть $h(a) = 0$. В этом случае, собственно, всё уже сказано: раз $\lim\limits_{u\to a} |f(x)| = \infty$, то надо доопределить $f(a)$ как ∞, и доопределённое таким образом отображение будет непрерывным в точке $u = a$. Поясним это подробнее. В различных частях $\mathbb{RP}^1$ мы используем, смотря по обстановке, координаты u и $v = \dfrac{1}{u}$. В окрестности точки $u = a$ годится координата u, а в окрестности той точки, которая выше обозначалась через ∞ — координата v (в самой этой точке $v = 0$). В этих терминах близкая к a точка с координатой $u \neq a$ переходит в точку с v-координатой

$$v(u) = \frac{1}{f(u)} = \frac{h(u)}{g(u)}.$$

Но дробь $\dfrac{h(u)}{g(u)}$ корректно определена также и в точке $u = a$, где она равна 0 (ведь $h(a) = 0$, а g и h не имеют общих корней). Раз $g(a) \neq 0$, то $g(u) \neq 0$ также и в некоторой окрестности точки a, и во всей этой окрестности дробь $\dfrac{h(u)}{g(u)}$ является непрерывной функцией u. Вот мы и доопределяем f в точке $u = a$ как ∞, т. е. как ту точку $\mathbb{RP}^1$, где $v = 0$. Тогда для любого $\varepsilon > 0$ найдётся такая окрестность точки $u = a$, в которой v-координата образа лежит в $(-\varepsilon, \varepsilon)$, что и означает непрерывность доопределённого отображения f в этой точке.

Осталось доопределить отображение f в точке ∞ и убедиться в непрерывности доопределённого отображения в этой точке. В окрестности последней мы используем координату v. Пусть степень многочлена g равна r, а степень h равна s, так что

$$g(u) = a_r u^r + \ldots + a_0, \quad h(u) = b_s u^s + \ldots + b_0, \quad a_r \neq 0, \ \ b_s \neq 0.$$

[52] Полюса рациональной функции $f = \dfrac{g}{h}$, где многочлены g и h не имеют общих корней (так что дробь несократима), — это, по определению, корни её знаменателя h. Когда x стремится к полюсу, $|f(x)| \to \infty$, а в самих полюсах $f(x)$ не определена.

При больших u мы представляем $g(u)$ и $h(u)$ в следующем виде:

$$g(u) = u^r\Bigl(a_r + a_{r-1}\frac{1}{u} + \ldots + a_1\frac{1}{u^{r-1}} + a_0\frac{1}{u^r}\Bigr) = u^r G(v),$$

где $G(v) = a_0v^r + a_1v^{r-1} + \ldots + a_{r-1}v + a_r$ $\left(\text{а } v = \dfrac{1}{u}\right)$, и аналогично

$$h(u) = u^s H(v), \quad \text{где } H(v) = b_0v^s + b_1v^{s-1} + \ldots + b_{s-1}v + b_s.$$

Поэтому $f(u) = u^{r-s}\dfrac{g(v)}{h(v)}$. Заметим, что при $v = 0$ дробь $\dfrac{g(v)}{h(v)}$ корректно определена и отлична от нуля — её числитель и знаменатель равны при этом ненулевым числам a_r и b_s. Если $r > s$, то $\lim\limits_{v\to 0} f(u) = 0$, а если $r = s$, то $\lim\limits_{v\to 0} f(u) = \dfrac{b_s}{a_r}$; в обоих случаях мы доопределяем $f(\infty)$, т. е. образ отображения f в точке $v = 0$, как соответствующий предел. Если $r < s$, то $\lim\limits_{u\to\infty} |f(u)| = \infty$. В этом случае мы пользуемся координатой v также и для образа отображения f. v-координата $f(u)$ равна $v^{s-r}\dfrac{b_s + b_{s-1}v + \ldots}{a_r + a_{r-1}v + \ldots}$, что стремится к 0 при $v \to 0$. Поэтому в данном случае мы доопределяем $f(\infty)$, т. е. образ f в точке $v = 0$, как точку с $v = 0$, т. е. полагаем $f(\infty) = \infty$. Во всех этих случаях при достаточной близости v к нулю (т. е. в достаточно малой окрестности точки $u = \infty$) $f(u)$ будет лежать в наперёд заданной окрестности точки $f(\infty)$ (продумайте это детальнее).

Замечание. Если f — многочлен, то выражение «степень f» имеет два смысла (оба совершенно стандартны): обычная степень многочлена f и степень отображения $f\colon \mathbb{RP}^1 \to \mathbb{RP}^1$. Чтобы избежать неясности, я обычно говорю не «степень f», а подробнее: «степень многочлена f», «степень отображения f».

Замечание. В $\mathbb{R}$ мы имеем обычные алгебраические операции — сложение и умножение. К сожалению, они не продолжаются до непрерывных операций в $\mathbb{RP}^1$. В элементарном курсе анализа объясняют, что в общем случае выражениям $\infty - \infty$ и $0 \cdot \infty$ невозможно придать смысл: если $x_n, y_n, w_n \to \infty$, $z_n \to 0$, то $x_n - y_n$ и z_nw_n могут стремиться к любому пределу, а могут и вовсе не иметь предела. («Хорошим» с топологической точки зрения является умножение точек $\mathbb{S}^1$ как комплексных чисел единичного модуля. Но оно как бы оторвано от проективных понятий. При желании читатель может получить для этой операции выражение в терминах аффинных координат и убедиться, что ничего хорошего оно не сулит.)

Поэтому возможность продолжить рациональную функцию до непрерывного отображения $\mathbb{RP}^1 \to \mathbb{RP}^1$ — это факт, который хотя и прост, но всё-таки потребовал разъяснений.

Если читателю тем не менее покажется, что всё сие более или менее тривиально, то я не буду возражать — мне как раз и хотелось бы, чтобы он пришёл к такому выводу. (С небольшим уточнением: тривиальным это является благодаря удачной системе понятий, выработка которой отнюдь не была тривиальным делом. Но мы-то ведь её не вырабатывали, а только осваиваем и используем, что куда проще.) А теперь извлечём некую пользу из этого толчения воды в ступе.

Доказательство правила Хованского

Доказательство правила Хованского сводится к трём утверждениям.

А. Число вещественных корней вещественного многочлена f равно степени отображения $\frac{f'}{f}\colon \mathbb{RP}^1 \to \mathbb{RP}^1$ со знаком минус.

Б. Если g — вещественный многочлен, то степень отображения $g\colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна введённой выше величине $j(g)$ (см. с. 87).

В. Если f — вещественная рациональная функция, представляющаяся в виде цепной дроби (34), то степень отображения $f\colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна

$$j(f_0) - j(f_1) + j(f_2) - \ldots + (-1)^k j(f_k).$$

Правило Хованского сразу следует из А и В. Согласно А, нам надо найти $-\operatorname{Ст}\frac{f'}{f}$, а ввиду В и (35) (т. е. того факта, что в цепной дроби для $\frac{f'}{f}$ отсутствует f_0),

$$\operatorname{Ст}\frac{f'}{f} = -j(f_1) + \ldots + (-1)^k j(f_k).$$

Б здесь не понадобилось. Но, во-первых, Б будет нужно для доказательства В, и, во-вторых, доказательство Б уместно предпослать доказательству А, потому что в обоих доказательствах используется последний раздел § 4, причём доказательство Б проще.

Итак, начнём с доказательства Б. Если степень многочлена g чётная — скажем, $2n$, так что $g(u) = a_{2n}u^{2n}$ + члены низших степеней, — то при всех достаточно больших $|u|$ — скажем, при $|u| > A$ — сумма членов низших степеней не превосходит по модулю $(1/2)|a_{2n}|u^{2n}$, и потому $g(u)$ имеет тот же знак, что и a_{2n}. Обозначим $\max\limits_{u\in[-A,A]}|g(u)|$ через M. Тогда получается, что при $a_{2n} > 0$ всюду $g(u) \geqslant -M$, а при $a_{2n} < 0$ всюду $g(u) \leqslant M$. Следовательно, $g(u)$ нигде на $\mathbb{R}$ не принимает некоторых значений, скажем, значения $b = -M-1$ в первом случае и $b = M+1$ во втором. В точке же $u = \infty$ значение g равно ∞, так что у b нет прообразов не только в $\mathbb{R}$, но и во всей проективной прямой. Следовательно, степень отображения g равна нулю. Таково же и $j(g)$ для многочлена g чётной степени. Если же степень g нечётная — скажем, $2n+1$, так что $g(u) = a_{2n+1}u^{2n+1}$ + члены низших степеней, то рассмотрим прообразы точки ∞. При конечных u значение $g(u)$ конечно, а $\lim\limits_{u\to\infty}|g(u)| = \infty$, поэтому $g^{-1}(\infty) = \infty$ (ясно ли, почему исключается возможность $g^{-1}(\infty) = \varnothing$?). Аналогично предыдущему, при достаточно больших по модулю u знак $g(u)$ совпадает со знаком $a_{2n+1}u^{2n+1}$, т. е. со знаком $a_{2n+1}u$ (ибо u^{2n+1} имеет тот же знак, что и u). В терминах v это означает, что если $a_{2n+1} > 0$, то при достаточно малом $\varepsilon > 0$ v-координата $\dfrac{1}{g(1/v)}$ (близкая к 0) имеет тот же знак, что и v, а если $a_{2n+1} < 0$, то знак v-координаты $\dfrac{1}{g(1/v)}$ противоположен знаку v. В первом случае образ точки с $v \neq 0$ лежит по ту же сторону от образа точки ∞ (для которой $v = 0$), по какую отображаемая точка лежит от ∞, поэтому точка ∞ является точкой подъёма. Во втором случае образ точки с $v \neq 0$ лежит по другую сторону от образа точки ∞, нежели отображаемая точка лежит от ∞, поэтому точка ∞ является точкой спуска. Согласно сказанному в конце § 4, в первом случае степень отображения g равна 1, а во втором случае она равна -1. Таково же и $j(g)$ для многочлена нечётной степени, и утверждение Б полностью доказано.

Доказательство утверждения А немногим сложнее. Как мы сейчас увидим, вещественные корни многочлена f — это прообразы точки ∞ при отображении $\dfrac{f'}{f}$, после чего всё сводится к проверке, что эти прообразы являются точками подъёма.

Когда мы интересуемся прообразами при каком-нибудь отображении, надо помнить, что в $\mathbb{RP}^1$, помимо чисел из $\mathbb{R}$, имеется ещё точка ∞, которая в принципе тоже может оказаться прообразом. Убедимся для начала, что сейчас на эту точку можно не обращать внимания: значение $\dfrac{f'}{f}$ в этой точке равно 0, а вовсе не ∞. Действительно, пусть

$$f(u) = a_n u^n + a_{n-1}u^{n-1} + \ldots + a_0.$$

Тогда, как известно,

$$f'(u) = na_n u^{n-1} + (n-1)a_{n-1}u^{n-2} + \ldots + a_1,$$

и так как в $\dfrac{f'}{f}$ степень числителя меньше степени знаменателя, то в точке ∞ отображению $\dfrac{f'}{f}$ приписывается значение 0. (Это случай $r < s$ выше. Выше, правда, отношение $\dfrac{g}{h}$ предполагалось несократимым, а сейчас не исключено, что f и f' могут иметь общие корни. Но при исследовании поведения $\dfrac{g}{h}$ в бесконечности несократимость $\dfrac{g}{h}$ не использовалась.)

Итак, прообразы ∞ при отображении $\dfrac{f'}{f}$ суть обычные числа. ∞ может быть образом обычного числа только если оно является корнем знаменателя. Проверим, что любой корень a многочлена f действительно переходит в ∞. Как известно, если a — корень кратности k, то $f(u)$ делится на $(u-a)^k$, а частное уже не обращается в нуль в точке a. Итак, $f(u) = (u-a)^k g(u)$ и $g(a) \neq 0$. Тогда

$$f'(u) = k(u-a)^{k-1}g(u) + (u-a)^k g'(u),$$
$$\frac{f'(u)}{f(u)} = \frac{k}{u-a} + \frac{g'(u)}{g(u)},$$

и, после перехода к координате v возле образа,

$$v = v(u) = \frac{(u-a)g(u)}{kg(u) + (u-a)g'(u)}.$$

В точке $u = a$ знаменатель равен $kg(a) \neq 0$, а числитель обращается в нуль. При малых $u-a$ как множитель $g(u)$ в числителе, так и знаменатель отличны от нуля и имеют тот же знак, что и $g(a)$. Поэтому $v(u)$ имеет при малых $u-a$ тот же знак, что и $u-a$.

При доказательстве Б возле прообраза и возле образа использовалась одна и та же координата. Теперь возле точки a используется координата u, а возле образа этой точки — координата v. Когда одна из них возрастает, другая убывает, поэтому направления их возрастания отвечают противоположным ориентациям $\mathbb{RP}^1$. Если считать положительной ориентацию, отвечающую возрастанию угловой координаты ψ, то из геометрического построения или из формулы $u = -\operatorname{tg}(\psi/2)$ (которую читатель должен был получить при решении упражнения 28) видно, что возрастанию ψ отвечает убывание u. Значит, возрастание v определяет на $\mathbb{RP}^1$ положительную ориентацию. Раз знак $v(u)$ совпадает со знаком $u-a$, то соответствующая точка $\frac{f'(u)}{f(u)}$ находится не по ту же сторону от ∞, по какую отображаемая точка с координатой u находится от точки $u = a$, а по другую. Получается, что корни многочлена f являются точками спуска для отображения $\frac{f'}{f}$. Согласно концу § 4, тогда степень этого отображения равна числу этих корней со знаком минус.

Наконец, докажем утверждение В. Для этого мы рассмотрим последовательно рациональные функции

$$g_0 = f_k, \quad g_1 = f_{k-1} + \frac{1}{f_k}, \;\ldots,\quad g_i = f_{k-i} + \cfrac{1}{f_{k-i+1} + \ddots + \cfrac{1}{f_k}}, \;\ldots$$
$$g_{k-1} = f_1 + \cfrac{1}{f_2 + \ddots + \cfrac{1}{f_k}}, \qquad g_k = f_0 + \cfrac{1}{f_1 + \cfrac{1}{f_2 + \ddots + \cfrac{1}{f_k}}}, \tag{38}$$

которые получаются при вычислении цепной дроби (35) «с конца» и которые можно более коротко определить так:

$$g_0 = f_k, \qquad g_i = f_{k-i} + \frac{1}{g_{i-1}} \text{ при } i = 1, \ldots, k. \tag{39}$$

Формально доказательство будет проведено индукцией по i (это как раз и означает, что мы последовательно рассматриваем $g_0, g_1, \ldots, g_k$). По индукции будут доказываться следующие два утверждения.

Г. При $i = 0, \ldots, k-1$ рациональная функция g_i является отношением многочленов $\frac{p_i(x)}{q_i(x)}$, где степень многочлена p_i больше степени многочлена q_i.

Д. Степень отображения $g_i \colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна

$$j(f_{k-i}) - j(f_{k-i+1}) + \ldots + (-1)^i j(f_k). \tag{40}$$

При $i = k$ утверждение Д совпадает с В. Утверждение Г в конечном счёте само по себе нам не нужно, но оно оказывается нужным для обоснования шага индукции (т. е. перехода от i к $i+1$) в доказательстве Д.

Доказательство Г несложно. При $i = 0$ $\ g_0 = f_k = \frac{f_k}{1}$. Степень многочлена f_k (как и всех f_i с $i > 0$) положительна, степень же постоянного многочлена 1 равна нулю. Тем самым Г доказано для $i = 0$. Допустим, что Г уже доказано для i, а $i + 1 < k$. Тогда

$$g_{k+1} = f_{k-i-1} + \frac{1}{g_k} = f_{k-i-1} + \frac{q_i}{p_i} = \frac{f_{k-i-1}p_i + q_i}{p_i}.$$

Степень многочлена $f_{k-i-1}p_i$ равна сумме степеней сомножителей и потому больше степени p_i (ведь $k-i-1 > 0$, так что степень многочлена f_{k-i-1} положительна), а последняя больше степени q_i. При добавлении к многочлену бóльшей степени m многочлена меньшей степени l степень суммы равна m (почему?). Получается, что

$$g_{i+1} = \frac{\text{многочлен степени} > \text{степени } p_i}{p_i}.$$

В доказательстве Д используются также следующие два утверждения.

Е. Если f — рациональная функция, не равная тождественно нулю, то степень отображения $\frac{1}{f} \colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна взятой со знаком минус степени отображения f.

Ж. Пусть g — многочлен, h — ненулевая рациональная функция, представимая в виде отношения многочленов $\frac{p(x)}{q(x)}$, где степень многочлена p

больше степени многочлена q, и пусть $f = g + \frac{1}{h}$. Тогда степень отображения $f\colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна $j(g) - \operatorname{Ст} h$.

Доказательство Е тоже совсем просто. Положим $g(x) = \frac{1}{x}$. Тогда отображение $\frac{1}{f} = g \circ f$, следовательно, $\operatorname{Ст}\frac{1}{f} = \operatorname{Ст} g \cdot \operatorname{Ст} f$, и остаётся только убедиться, что степень отображения $g\colon \mathbb{RP}^1 \to \mathbb{RP}^1$ равна -1. Когда u пробегает отрезок $(0, \infty)$, возрастая, то $\frac{1}{u}$ пробегает тот же отрезок, убывая. Больше нигде в $\mathbb{RP}^1 = \mathbb{R} \cup \{\infty\}$ — ни на полуоси $(-\infty, 0)$, ни в точке 0, ни в точке ∞, — функция g не принимает значений из $(0, \infty)$. Значит, для любого $b \in (0, \infty)$ прообраз $g^{-1}(b)$ сводится к единственной точке $a = \frac{1}{b} \in (0, \infty)$. При этом

$$g(a - \varepsilon, a) = \left(\frac{1}{a}, \frac{1}{a - \varepsilon}\right) > \frac{1}{a} = b, \qquad g(a, a + \varepsilon) = \left(\frac{1}{a + \varepsilon}, \frac{1}{a}\right) < b,$$

так что a есть точка спуска. Итак, для g разность

$$\text{число точек подъёма} - \text{число точек спуска} = 0 - 1 = -1.$$

Приступаем к доказательству утверждения Ж. Рассмотрим прообраз $f^{-1}(\infty)$. Содержит ли он точку ∞? Поскольку степень многочлена q меньше степени многочлена p, то в этой точке $\frac{q}{p} = 0$, а в некоторой её окрестности $\left|\frac{q}{p}\right| < 1$. Если степень n многочлена g равна нулю, т. е. он равен некоторой константе c, то в некоторой окрестности точки ∞ всюду $\left|g + \frac{q}{p}\right| < |c| + 1$, так что $\infty \notin f^{-1}(\infty)$. Если же эта степень равна $n > 0$, то $g(\infty) = \infty$ и $f(\infty) = \infty$. (Здесь не надо даже переходить к координате $v = \frac{1}{u}$ — работает очевидное рассуждение в духе элементарного курса анализа: $|g(x)| \to \infty$ при $|x| \to \infty$, $\left|\frac{q}{p}\right| < 1$, значит, и $|f(x)| \to \infty$. В терминах непрерывного отображения $f\colon \mathbb{R} \cup \{\infty\} \to \mathbb{R} \cup \{\infty\}$ получается, что $f(x) \to \infty$ при $x \to \infty$. Значит, $f(\infty) = \infty$.) Итак, в этом случае $\infty \in f^{-1}(\infty)$. Проверим, что, в зависимости от того, будет ли для g точка ∞ точкой подъёма, спуска или локального экстремума, она будет такой же точкой и для f.

Если степень n многочлена g чётна и старший коэффициент $a_n > 0$ (соответственно, $a_n < 0$), то для любого $\varepsilon > 0$ найдётся такое $\delta > 0$, что при $|x| > \frac{1}{\delta}$ будет $g(x) > \frac{1}{\varepsilon}$, а тогда $f(x) = g(x) + \frac{q(x)}{p(x)} > \frac{1}{\varepsilon} - 1$ (соответственно, $f(x) < -\frac{1}{\varepsilon} + 1$). Поэтому функция f не принимает в некоторой окрестности точки ∞ отрицательных (соответственно, положительных) значений; выходит, что образ этой окрестности при отображении f лежит по одну сторону от ∞. Стало быть, в этом случае ∞ является точкой локального экстремума для f (как и для g). Если n нечётно и $a_n > 0$ (соответственно, $a_n < 0$), то аналогично находим, что

$$g\Big(-\infty, -\frac{1}{\delta}\Big) \subset \Big(-\infty, -\frac{1}{\varepsilon}\Big), \qquad g\Big(\frac{1}{\delta}, \infty\Big) \subset \Big(\frac{1}{\varepsilon}, \infty\Big)$$

(соответственно,

$$g\Big(-\infty, -\frac{1}{\delta}\Big) \subset \Big(\frac{1}{\varepsilon}, \infty\Big), \qquad g\Big(\frac{1}{\delta}, \infty\Big) \subset \Big(-\infty, -\frac{1}{\varepsilon}\Big).)$$

Значит, в этом случае ∞ является точкой подъёма (соответственно, спуска) для f, как и для g.

Теперь займёмся прообразами ∞, являющимися обычными числами. Для любого такого прообраза a должно быть $h(a) = 0$ — иначе $g(a) + \dfrac{1}{h(a)}$ было бы обыкновенным («конечным») числом. Идея состоит в том, что возле точки a, где $h(a) = 0$, первое слагаемое в сумме $g(x) + \dfrac{1}{h(x)}$ слабо меняется сравнительно со вторым, поэтому поведение суммы определяется вторым слагаемым. Покажем, что a является

а) точкой подъёма для f, если a является точкой спуска для h;

б) точкой спуска для f, если a является точкой подъёма для h;

в) точкой локального экстремума для f, если a является точкой локального экстремума для h.

Дробь $\dfrac{p}{q}$, представляющую h, мы вправе считать несократимой (почему?). В таком случае, если $h(a) = \dfrac{p(a)}{q(a)} = 0$, то $p(a) = 0$, $q(a) \neq 0$. Если p имеет в a нуль кратности k, то $p(u) = (u-a)^k r(u)$, где r — некоторый многочлен и $r(a) \neq 0$. Значит,

$$f(u) = g(u) + \frac{q(u)}{p(u)} = \frac{(u-a)^k rg + q}{(u-a)^k r}.$$

Отображение f переводит точку с координатой u в точку с координатой $U = f(u)$. Так как при $u \approx a$ образ близок к ∞, то естественно перейти в образе к координате $V = V(u) = \dfrac{1}{U} = \dfrac{1}{f(u)}$. Тогда

$$V = \frac{(u-a)^k r}{(u-a)^k rg + q}.$$

При $u = a$ знаменатель $= q(a) \neq 0$, а числитель $= 0$, и $V = 0$, как это и должно быть для v-координаты точки ∞. При малом $(u-a)$ знак $V(u)$

совпадает со знаком $(u-a)^k$ или противоположен ему, в зависимости от знака $\dfrac{r(a)}{q(a)}$.

Если k чётно, то V имеет один и тот же знак и при $u < a$, и при $u > a$; следовательно, a — точка локального экстремума для f. Но для $h(u) = \dfrac{(u-a)^k r(u)}{q(u)}$ она тоже является точкой локального экстремума.

Если же k нечётно, то знак $(u-a)^k$ меняется при прохождении u через a, а знак $\dfrac{r(u)}{(u-a)^k r(u) g(u) + q(u)}$ возле a постоянен и совпадает со знаком $\dfrac{r(u)}{q(u)}$, т. е. со знаком $\dfrac{r(a)}{q(a)}$. Итак, в этом случае

$$V(a-\varepsilon,\, a) < 0 \quad \text{и} \quad V(a,\, a+\varepsilon) > 0, \qquad \text{если } \frac{r(a)}{q(a)} > 0, \tag{41}$$

$$V(a-\varepsilon,\, a) > 0 \quad \text{и} \quad V(a,\, a+\varepsilon) < 0, \qquad \text{если } \frac{r(a)}{q(a)} < 0. \tag{42}$$

Но надо учесть ещё, что возле a мы используем координату u, возле ∞ — координату v, а направления возрастания v и u противоположны. Поэтому в случае (41) a является точкой спуска для f, а в случае (42) — точкой подъёма. Что же до h, то $h(u)$ при $\dfrac{r(a)}{q(a)} > 0$ имеет возле точки a тот же знак, что и $(u-a)$, а при $\dfrac{r(a)}{q(a)} < 0$ знак $h(u)$ возле точки a противоположен знаку $(u-a)$. Но для h возле точек a и $h(a) = 0$ мы используем одну и ту же координату u. Поэтому вывод о типе прообраза a точки 0 для отображения h противоположен сделанному выше выводу о типе прообраза a точки ∞ для отображения f.

Окончательно получаем следующее. Если в прообразе $h^{-1}(0)$ (который целиком содержится в $\mathbb{R}$ — ведь $h(\infty) = \infty$, а не 0) имеется k точек подъёма и l точек спуска для h, то $\operatorname{Ст} h = k - l$. Для f те же точки являются, наоборот, точками спуска и подъёма, и, кроме того, точка ∞ является точкой подъёма, спуска, локального экстремума или вообще не принадлежит $f^{-1}(\infty)$, смотря по тому, прообразом какого типа она является для g. Стало быть, $\operatorname{Ст} f = j(g) - \operatorname{Ст} h$, и утверждение Ж доказано.

После всего сказанного утверждение Д легко доказывается по индукции. Для g_0 нечего доказывать. Пусть Д доказано для g_i, т. е. $\operatorname{Ст} g_i$ даётся выражением (40). Если $i = k$, то опять-таки нечего доказывать (в этом случае никакого g_{i+1} у нас нет и о нём ничего не утверждается). Содержателен переход от i к $i+1$ при $i < k$. Согласно Г, g_i имеет вид $g_i = \dfrac{p_i(x)}{q_i(x)}$, где степень многочлена p_i больше степени многочлена q_i. Согласно Ж, тогда степень отображения g_{i+1} равна $j(f_{k-i-1}) - \operatorname{Ст} g_i$. Подставив сюда доказанное (по предположению индукции) выражение (40) для степени отображения g_i, получим выражение для g_{i+1}, которое отличается от выражения для $\operatorname{Ст} g_i$ только заменой i на $i+1$.

§ 7. Критерий Эйленберга

Начиная отсюда, мы занимаемся доказательством теоремы Жордана. Пока мы имеем дело со сравнительно простыми кривыми или с небольшим числом конкретных кривых, имеющих более сложную форму, но всё же локально устроенных достаточно просто (скажем, кусочно гладких), — до тех пор в теореме Жордана нет необходимости, ибо в существовании внутренних точек можно убедиться эмпирически, путём созерцания или штриховки. Этим объясняется, почему теорема Жордана была сформулирована только во второй половине XIX века. Между прочим, доказательство самого К. Жордана было неполным, но он, во всяком случае, понял, что такие вещи нужно доказывать. Первые строгие доказательства были найдены в начале XX века, см. исторические замечания в [1].

В одном из первых доказательств теоремы Жордана, данном Брауэром, использовался «технический аппарат» так называемой алгебраической топологии, который Брауэр как раз и разрабатывал в то время. При использовании подобного аппарата (особенно если его развить чуть дальше, чем это было тогда у самого Брауэра) доказательство не требует больших усилий и получается коротким. Но ввиду простоты формулировки теоремы Жордана представляют интерес и элементарные доказательства, доступные тем, кто не знаком с упомянутым «техническим аппаратом». Согласно [1], два таких доказательства были опубликованы ещё до работы Брауэра, но и после неё время от времени публиковались новые элементарные доказательства. Однако, так как они не используют поддержки, доставляемой упомянутым аппаратом, то они оказываются более длинными и громоздкими [53)], а потому снова и снова возникает желание придумать новое элементарное доказательство, которое было бы «лучше» (короче, прозрачнее) прежних.

Если говорить о коротком и прозрачном доказательстве, не использующем никаких вспомогательных средств, то, мне кажется, особых надежд на него питать не следует — накопленный опыт свидетельствует, что совсем «голыми руками» теорему Жордана доказать непросто. Но можно использовать «технический аппарат», который гораздо проще, чем у Брауэра — изложенную в § 4 гомотопическую теорию отображений окружности в окружность. Общая идея моего подхода принадлежит С. Эйленбергу [2];

[53)] Порой слишком короткие элементарные доказательства содержат пробелы.

на русском языке он изложен в [3]. Но буквально я следую Эйленбергу только в настоящем параграфе, где приводится некоторое общее условие, характеризующее точки, расположенные внутри замкнутой кривой. Мне кажется, здесь у Эйленберга всё так хорошо, что улучшить его невозможно. А вот дальше в [2], [3] доказательство распадается на большое число мелких шагов, и при всей их простоте как-то теряется общая перспектива[54)]. Я предлагаю другое рассуждение, в котором главное содержится в одном большом шаге. Он, правда, тоже расчленяется на несколько частей, но их меньше и, что важнее, общее направление мысли не ускользает из виду (так, по крайней мере, мне кажется).

Я уже упоминал мельком, что тривиальность доказательства теоремы Жордана в простых случаях связана с тем, что точки, лежащие внутри нашей кривой, можно легко охарактеризовать при помощи простого аналитического или геометрического условия, формулировка которого никак не зависит от теоремы Жордана (скажем, в случае окружности это условие состоит в том, что расстояние точки до центра окружности меньше её радиуса, а для периметра квадрата, вершины которого имеют декартовы координаты (0, 0), (1, 0), (1, 1), (0, 1), — это условие $0 < x, y < 1$). А после этого в рассматриваемых случаях действительно становится совершенно ясно, что если точка $z(t)$, непрерывно зависящая от t, при одном t удовлетворяет упомянутому условию (скажем, отстоит от центра окружности меньше чем на её радиус), а при другом t этому условию не удовлетворяет, то при каком-то промежуточном значении t она должна лежать на данной кривой (в примере с окружностью это явствует из того, что расстояние до центра само будет непрерывной функцией t). То есть из той части плоскости, где выполняется данное условие, нельзя непрерывно перейти в ту часть, где оно не выполняется, так что действительно $\Pi \setminus L$ состоит по крайней мере из двух частей. В существовании точек, удовлетворяющих этому условию, тоже легко убедиться. Значит, в таких случаях дополнение к замкнутой кривой состоит по крайней мере из двух связных частей; в том, что их ровно две, тоже нетрудно убедиться.

Беда в том, что в различных простых случаях соответствующие условия оказываются различными, и не видно, что могло бы их заменить в общем случае. Критерий Эйленберга, которому посвящён настоящий параграф,

---

[54)] Непосредственно это относится более к [3], чем к [2]. Дело в том, что в [2] теорема Жордана — только один из результатов топологии плоскости, доказываемый сравнительно поздно, причём при доказательстве привлекаются некоторые из ранее полученных результатов. При таком содержании [2] там и не должна была преследоваться цель написать рассуждение, которое, отправляясь от исходных понятий и соответствующих результатов, прямо вело бы к теореме Жордана. Но если попытаться выделить из [2] только то, что позволяет доказать теорему Жордана, то получится более или менее то же самое, что написано в [3].

является таким условием для общего случая. Он совсем не похож по своей формулировке на очевидные условия, сразу приходящие в голову для окружности, квадрата и т. п.

В этом параграфе K — компактное подмножество $\mathbb{R}^2$. Через f_a, где точка $a \in \mathbb{R}^2 \setminus K$, всё время обозначается следующее отображение $K \to \mathbb{S}^1$:

$$f_a(z) = \frac{z-a}{|z-a|} \qquad (z \in K).$$

Теорема Эйленберга. *Точки $a, b \in \mathbb{R}^2 \setminus K$ тогда и только тогда лежат в одной компоненте связности множества $\mathbb{R}^2 \setminus K$, когда отображения f_a и f_b гомотопны.*

Например, если $K = \mathbb{S}^1$, то для всех точек a внешней области, которая в данном случае есть $\{x;\, |x| > 1\}$, f_a несущественно, тогда как для всех точек другой компоненты связности $\mathbb{C} \setminus \mathbb{S}^1$, которая есть $\{x;\, |x| < 1\}$, f_a является отображением $\mathbb{S}^1 \to \mathbb{S}^1$ степени 1.

Доказательство. Необходимость очевидна: если a, b лежат в одной компоненте связности U дополнения к K в $\mathbb{C}$, то их можно соединить в U путём $\{a(t)\}$, $a(0) = a$, $a(1) = b$, и тогда $\{f_{a(t)}\}$ — гомотопия, соединяющая f_a и f_b.

Переходим к достаточности. Пусть U, V — две различные компоненты связности $\mathbb{C} \setminus K$, $a \in U$, $b \in V$ и $f_a \sim f_b$. Мы вправе считать, что U не является внешней областью для K (в противном случае мы поменяли бы ролями a, U и b, V). Так как непрерывное отображение $f_a/f_b\colon K \to \mathbb{S}^1$ несущественно, то у него имеется угловая функция $\varphi\colon K \to \mathbb{R}$, так что $f_a/f_b = e^{i\varphi}$, $f_a = f_b e^{i\varphi}$. Функцию φ можно продолжить до непрерывной функции $\Phi\colon \mathbb{C} \to \mathbb{R}$. Положим для всех $w \in \mathbb{C}$

$$f(w) = \begin{cases} \dfrac{w-a}{|w-a|} & \text{при } w \in \mathbb{C} \setminus U, \\ e^{i\Phi(w)} \dfrac{w-b}{|w-b|} & \text{при } w \in \operatorname{clos} U. \end{cases}$$

«Области действия» первой и второй строчек суть замкнутые множества, пересекающиеся по границе $\operatorname{Fr} U$. Очевидно, $\operatorname{Fr} U \subset K$, ибо каждая точка x из открытого множества $\mathbb{C} \setminus K$ является внутренней точкой одной из компонент связности этого множества (скажем, компоненты W) и потому не может быть граничной точкой для U (когда $W = U$, речь идёт просто о различии между внутренними и граничными точками U, а когда $U \neq W$, из $x \in W \cap \operatorname{Fr} U$ следовало бы, что $U \cap W \neq \varnothing$ — ведь сколь угодно близко к $x \in \operatorname{Fr} U$ имеются точки U, а все достаточно близкие к $x \in W$ точки принадлежат W). Поэтому в точках $\operatorname{Fr} U$

$$e^{i\Phi(w)} \frac{w-b}{|w-b|} = e^{i\varphi(w)} \frac{w-b}{|w-b|} = \frac{w-a}{|w-a|},$$

т. е. первая и вторая строчки в определении f приписывают этим точкам одинаковые значения $f(w)$. Стало быть, отображение $f\colon \mathbb{C} \to \mathbb{S}^1$ корректно определено. Оно непрерывно, ибо в обеих строчках определения фигурируют непрерывные функции.

Теперь нагляднее рассматривать f как непрерывное векторное поле [55]. Оно не имеет нулей, ибо всюду $|f| = 1$. При достаточно большом r окружность $C = \{w;\ |w - a| = r\}$ целиком содержится во внешней области и тем более в $\mathbb{C} \setminus U$, поэтому в её точках $f(w) = \dfrac{w-a}{|w-a|}$. Для такого поля $\operatorname{Вр}_C f = 1$, а потому оно должно обращаться в нуль где-то внутри C.

Из теоремы Эйленберга следует, что простая дуга J (т. е. гомеоморфный образ I) не разбивает плоскости. Действительно, любое непрерывное отображение $f\colon J \to \mathbb{S}^1$ несущественно.

Другое простое следствие состоит в том, что если Σ^1 — топологическая окружность на плоскости $\mathbb{C}$, то Σ^1 является границей любой компоненты связности U своего дополнения $\mathbb{C} \setminus U$. (При этом мы пока не знаем, сколько имеется таких компонент.) При доказательстве теоремы Эйленберга мы уже видели, что $\operatorname{Fr} U \subset \Sigma^1$. Там речь шла о более общем случае дополнения к компактному множеству K, при этом $\operatorname{Fr} U$ вполне могла быть только частью K. Теперь же надо доказать, что $\Sigma^1 \subset \operatorname{Fr} U$.

Пусть $f\colon \mathbb{S}^1 \to \Sigma^1$ — гомеоморфизм. Допустим, что имеется точка $x \in \mathbb{S}^1 \setminus \operatorname{Fr} U$. Так как множество $\operatorname{Fr} U$ — замкнутое, у x имеется окрестность W, не пересекающаяся с $\operatorname{Fr} U$. Прообраз $f^{-1}(W \cap \Sigma^1)$ — окрестность точки $f^{-1}(x)$ на $\mathbb{S}^1$. Эта окрестность должна содержать некоторую открытую дугу, на которой лежит $f^{-1}(x)$. Дополнение J к этой дуге на $\mathbb{S}^1$ — замкнутая дуга (т. е. дуга, содержащая свои концы). Она гомеоморфна I, так что её образ $f(J)$ при отображении f является простой дугой на плоскости $\mathbb{C}$. $f(J)$ не разбивает $\mathbb{C}$, так что любые две точки его дополнения $\mathbb{C} \setminus f(J)$ можно соединить путём, не пересекающим $f(J)$. С другой стороны, $x \notin \operatorname{clos} U$ (ведь $x \notin U$ и $x \notin \operatorname{Fr} U$), поэтому путь, соединяющий x с какой-нибудь точкой $y \in U$, должен пересекать $\operatorname{Fr} U$ и тем более $f(J)$; а ведь обе эти точки лежат в $\mathbb{C} \setminus f(J)$ (почему?).

[55] Обычно мы обозначали векторы иначе, чем точки или комплексные числа, — с помощью стрелок или полужирного шрифта. Но, после сказанного по поводу наших отождествлений

функция с векторными значениями = векторное поле,

вектор = точка = пара чисел (x, y) = комплексное число,

хотя они и являются временными, мы позволим себе не менять обозначений, когда мы меняем точку зрения, рассматривая функцию с комплексными значениями как векторное поле или обратно.

§ 8. Топологическая окружность разбивает плоскость

Теорема. *Пусть A — связное подмножество*[56] *какого-нибудь $\mathbb{R}^n$, а $F\colon \mathbb{S}^1 \times A \to \mathbb{S}^1$ — непрерывное отображение. Тогда F можно представить в виде* («*комплексная*» *запись, подразумевающая использование умножения и единицы в $\mathbb{C}$, а также комплексной экспоненты*):

$$F(z, w) = F(z, w_0)F(1, w)e^{i\Phi(z,w)},$$

где w_0 — произвольная фиксированная точка A, а $\Phi\colon \mathbb{S}^1 \times A \to \mathbb{R}$ — непрерывная вещественная функция.

Эту теорему можно прокомментировать с двух точек зрения. С одной стороны, можно говорить о семействе непрерывных отображений

$$F_w\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad F_w(z) = F(z, w),$$

параметризованном посредством $w \in A$ (тогда как раньше мы имели дело с семействами, параметризованными посредством $t \in I$). Теорема как бы сравнивает различные F_w с F_{w_0}. С другой стороны, если имеются отображения в окружность каких-то двух множеств

$$f_1\colon A_1 \to \mathbb{S}^1, \qquad f_2\colon A_2 \to \mathbb{S}^1,$$

то их можно перемножить, перемножая значения этих отображений в точках $w_1 \in A_1$, $w_2 \in A_2$ как обычные комплексные числа:

$$A_1 \times A_2 \to \mathbb{S}^1, \qquad (w_1, w_2) \to f_1(w_1)f_2(w_2).$$

Теорема утверждает, что в нашем случае отображение F гомотопно произведению отображений

$$\mathbb{S}^1 \to \mathbb{S}^1, \qquad z \mapsto F(z, w_0),$$

и

$$A \to \mathbb{S}^1, \qquad w \mapsto F(1, w).$$

Утверждение, совершенно аналогичное последнему, имеет место для непрерывных отображений произведения двух связных компактных множеств[57] A_1, A_2 в окружность $\mathbb{S}^1$: всякое такое отображение гомотопно произведению двух непрерывных отображений $A_1 \to \mathbb{S}^1$, $A_2 \to \mathbb{S}^1$. (Это фактически известно по меньшей мере с 30-х гг., хотя я не знаю, кто и когда первым сформулировал явно данное утверждение.) Условие связности здесь существенно: если только имеются два негомотопных друг другу

[56] Более общо, A может быть любым связным метрическим или топологическим пространством.

[57] Более общо, двух связных компактных топологических пространств.

отображения $f', f''\colon A_1 \to \mathbb{S}^1$, а A_2 несвязно и представляется в виде объединения $A_2' \cup A_2''$ двух непересекающихся открыто-замкнутых в A_2 подмножеств, то можно положить

$$F(w_1, w_2) = \begin{cases} f'(w_1) & \text{при } w_2 \in A_2', \\ f''(w_1) & \text{при } w_2 \in A_2''. \end{cases}$$

Тогда в семействе отображений $F_{w_2}\colon w_1 \mapsto F(w_1, w_2)$ (получающихся при различных w_2) будут иметься негомотопные друг другу отображения. Но если бы F было гомотопно отображению вида $(w_1, w_2) \mapsto f_1(w_1)f_2(w_2)$, то из этой гомотопии понятным образом получалась бы и гомотопия каждого из отображений F_{w_2} в отображение $w_1 \mapsto f_1(w_1)f_2(w_2)$, а последнее при каждом w_2 гомотопно отображению f_1 («передвигаем» число $f_2(w_2)$ по $\mathbb{S}^1$ в положение 1; другое дело, что гомотопию, вообще говоря, нельзя сделать непрерывной по w_2, но этого сейчас и не нужно). Можно показать, что условие компактности тоже существенно, хотя достаточно компактности одного из сомножителей. Достаточность (вернее, эквивалентное ей утверждение) доказана в [2] (утверждение (8) из § 2 первой части) [58]. В нашей теореме мы как раз и имеем дело с таким случаем, когда не предполагается компактности одного из сомножителей (A), зато второй сомножитель не только компактен, но является простой конкретной фигурой ($\mathbb{S}^1$). Вероятно, теорема выглядит естественнее, если воспринимать её на фоне сказанного, но мы ограничимся только тем, что нам нужно.

Доказательство. Рассмотрим непрерывное отображение

$$G\colon \mathbb{S}^1 \times A \to \mathbb{S}^1, \qquad G(z, w) = \frac{F(z, w)F(1, w_0)}{F(1, w)F(z, w_0)}.$$

Для него $G(1, w) = G(z, w_0) = 1$ при всех $z \in \mathbb{S}^1$, $w \in A$. В этом частном случае теорема утверждает, что G можно представить в виде $e^{i\Psi(z,w)}$ с некоторой непрерывной функцией $\Psi\colon \mathbb{S}^1 \times A \to \mathbb{R}$. Обратно, если в этом случае теорема будет доказана, то для F получится

$$F(z, w) = \frac{F(1, w)F(z, w_0)G(z, w)}{F(1, w_0)} = F(1, w)F(z, w_0)e^{i(\Psi(z,w)-\alpha)},$$

где α — один из аргументов числа $F(1, w_0)$. Итак, мы можем ограничиться доказательством теоремы для частного случая отображения G.

Рассмотрим семейство отображений

$$g_t\colon A \to \mathbb{S}^1, \qquad g_t(w) = G(e^{2\pi i t}, w) \qquad (w \in A,\ t \in I).$$

[58] Но там нет ничего похожего на излагаемое ниже построение поля R, использующее это утверждение.

При $t = 0$ всё A отображается в 1. Значит, имеются такие угловые функции $\chi_t \colon A \to \mathbb{R}$, что $\chi_t(w)$ непрерывно по (w, t) и $\chi_0 = 0$. Что можно сказать о χ_1?

Раз $e^{i\chi_1(w)} = g(1, w) = 1$, то $\frac{1}{2\pi}\chi_1(w)$ принимает целые значения. Но $\chi_1(w)$ — непрерывная функция от w, а A связно. Следовательно, $\chi_1(w) = 2\pi l$ с некоторым целым l, не зависящим от w. Чтобы определить l, положим $w = w_0$. Функция $t \mapsto \chi_t(w_0)$ — угловая функция для отображения

$$t \mapsto g_t(w_0) = G(e^{2\pi i t}, w_0) = 1.$$

При $t = 0$ эта угловая функция равна нулю (как и вообще все $\chi_0(w)$). Значит, и при всех t она равна нулю ($t \mapsto \frac{1}{2\pi}\chi_t(w_0)$ — угловая функция на I с целыми значениями). В частности, $\chi_1(w_0) = 0$, откуда $l = 0$, т. е. $\chi_1(w) = 0$ при всех w.

А теперь можно положить $\Psi(e^{2\pi i t}, w) = \chi_t(w)$. Определение корректно, т. к. хотя $1 = e^{2\pi i t}$ и при $t = 0$, и при $t = 1$, но оба варианта дают одно и то же. Очевидна непрерывность этой функции в точках $(\mathbb{S}^1 \setminus \{1\}) \times A$. Если же $(z_k, w_k) \to (1, w)$, т. е. $z_k \to 1$, $w_k \to w$, то $z_k = e^{2\pi i t_k}$, где $t_k \in I$, так что при достаточно больших k будет либо $0 \leqslant z_k < \delta$, либо $1 - \delta < z_k \leqslant 1$, где δ — наперёд заданное сколь угодно малое положительное число. В обоих случаях при достаточной малости δ будет обеспечена близость $\chi_{t_k}(w_k)$ к $\chi_0(w)$ или $\chi_1(w)$, а это одно и то же число, которому как раз и равно $\Psi(1, w)$.

Теорема. *Топологическая окружность Σ^1 на плоскости $\mathbb{C}$ разбивает последнюю.*

Доказательство. Пусть Σ^1 является образом $\mathbb{S}^1$ при гомеоморфизме $j \colon \mathbb{S}^1 \to \Sigma^1$ и V — внешняя область для Σ^1. Мы построим такое непрерывное отображение $R \colon V \cup \Sigma^1 \to \mathbb{S}^1$, что для всех достаточно далёких x будет $R(x) = \frac{j(1) - x}{|j(1) - x|}$. Рассматривая R как векторное поле, получим, что на окружности $C = \{x;\ |x - j(1)| = r\}$ с достаточно большим r будет $\mathrm{Вр}_C R = 1$. Следовательно, $\mathbb{C} \setminus \Sigma^1 \neq V$, ибо в противном случае было бы $\mathbb{C} = V \cup \Sigma^1$, так что R было бы непрерывным полем на всём $\mathbb{C}$, нигде не равным нулю; но раз $\mathrm{Вр}_C R = 1$, то оно должно иметь нуль где-то внутри C.

На Σ^1 рассмотрим векторное поле $p(z) = \frac{j(-j^{-1}(z)) - z}{|j(-j^{-1}(z)) - z|}$. Когда $\Sigma^1 = \mathbb{S}^1$ и j — тождественное отображение, $j(e^{i\varphi}) = e^{i\varphi}$ при всех φ, то $-j^{-1}(z) = -z$ — точка, диаметрально противоположная z, $j(-j^{-1}(z))$ — та же самая точка и $p(z) = -z$. Рассматривая p как вектор, получаем векторное поле, векторы которого имеют начало z и конец 0. Для этого поля

$\operatorname{Bp}_{\mathbb{S}^1} p = 1$, поэтому при любом его непрерывном продолжении на всю плоскость $\mathbb{C}$ у него будут где-то нули. Во внешнюю же область окружности $\mathbb{S}^1$ его, конечно, легко продолжить как поле без нулей (как поле $z \mapsto -z$ или, если желать, чтобы векторы поля имели единичную длину, $z \mapsto \frac{z}{|z|}$). В общем случае наше p является некоторым обобщением этого поля: $j(-j^{-1}(z))$ есть точка Σ^1, как бы диаметрально противоположная точке z (см. рис. 37). Для него мы докажем то же, что сказано о предыдущем примере, т. е. что p продолжается во внешнюю область V как поле без нулей, а на всю $\mathbb{C}$ так не продолжается, откуда ясно, что $V \cup \Sigma^1 \neq \mathbb{C}$.

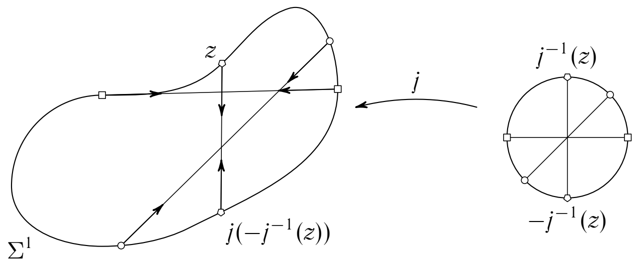


Рис. 37

Сперва мы продолжим p до некоторого непрерывного поля P в некотором открытом множестве U, содержащем Σ^1, и получим некоторое представление для P в $U \cap V$.

Непрерывная функция

$$\Sigma^1 \to \mathbb{C}, \qquad z \mapsto -j^{-1}(z)$$

может быть продолжена до непрерывной функции $G\colon \mathbb{C} \to \mathbb{C}$. Положим $U = G^{-1}(\mathbb{C} \setminus \{0\})$. Это — открытое множество, содержащее Σ^1 (почему?). На нём корректно определено и непрерывно отображение

$$H\colon U \to \mathbb{S}^1, \qquad H(u) = \frac{G(u)}{|G(u)|},$$

которое на Σ^1 совпадает с отображением $z \mapsto -j^{-1}(z)$ (см. рис. 38). Поэтому на U определено и непрерывно отображение

$$P\colon U \to \mathbb{S}^1, \qquad P(u) = \frac{j(H(u)) - u}{|j(H(u)) - u|}.$$

Рассмотрим непрерывное отображение

$$F\colon \mathbb{S}^1 \times V \to \mathbb{S}^1, \qquad F(e^{i\varphi}, v) = \frac{j(e^{i\varphi}) - v}{|j(e^{i\varphi}) - v|}$$

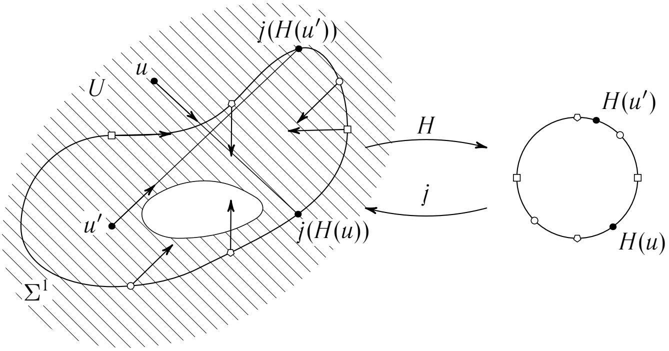


Рис. 38

(знаменатель всюду отличен от 0, потому что $j(e^{i\varphi}) \in \Sigma^1$, $v \notin \Sigma^1$). Тогда

$$P(v) = F(H(v), v) \quad \text{при} \quad v \in U \cap V.$$

С другой стороны, согласно последней теореме, зафиксировав какую-нибудь точку $v_0 \in V$, имеем

$$F(e^{i\varphi}, v) = F(e^{i\varphi}, v_0)F(1, v)e^{i\Phi(e^{i\varphi}, v)},$$

где Φ — непрерывная функция своих аргументов с вещественными значениями. В частности,

$$P(v) = F(H(v), v_0)F(1, v)e^{i\Phi(H(v), v)} \quad \text{при} \quad v \in U \cap V.$$

Займёмся первым множителем:

$$F(H(v), v_0) = \frac{j(H(v)) - v_0}{|j(H(v)) - v_0|} = q(H(v)),$$

где q есть отображение

$$q\colon \mathbb{S}^1 \to \mathbb{S}^1, \qquad q(e^{i\varphi}) = \frac{j(e^{i\varphi}) - v_0}{|j(e^{i\varphi}) - v_0|}.$$

Если взять точку v_0 достаточно далеко, то из неё вся топологическая окружность Σ^1 будет видна под малым углом, и тогда ясно, что степень отображения q равна нулю. Следовательно, оно представимо в виде $q(e^{i\varphi}) = e^{i\psi(e^{i\varphi})}$ с некоторой непрерывной функцией $\psi\colon \mathbb{S}^1 \to \mathbb{R}$. Тогда

$$F(H(v), v_0) = e^{i\psi(H(v))},$$

а значит,

$$P(v) = \frac{j(1) - v}{|j(1) - v|}e^{i(\psi(H(v)) + \Phi(H(v), v))} = \frac{j(1) - v}{|j(1) - v|}e^{i\varphi(v)}, \quad v \in U \cap V,$$

где $\varphi(v) = \psi(H(v)) + \Phi(H(v), v))$ — непрерывная функция на $U \cap V$ со значениями в $\mathbb{R}$.

Функция $\xi\colon \Sigma^1 \to \mathbb{R}$, тождественно равная 1, допускает непрерывное продолжение Ξ на всё $\mathbb{R}^n$, равное нулю всюду вне U. Я хочу чуть бо́льшего: чтобы она равнялась 1 всюду на некотором открытом множестве W_1, содержащем Σ^1, и 0 вне некоторого бо́льшего множества W_2, замыкание которого лежит в U. Введём обозначения (см. рис. 39):

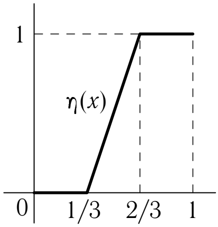


Рис. 39

$$\eta(x) = \begin{cases} 0 & \text{при } x \leqslant \dfrac{1}{3}, \\ 3\Big(x - \dfrac{1}{3}\Big) & \text{при } x \in \Big[\dfrac{1}{3}, \dfrac{2}{3}\Big], \\ 1 & \text{при } x \geqslant \dfrac{2}{3}, \end{cases}$$

$$\zeta(x) = \eta(\Xi(x)),\ W_1 = \Xi^{-1}\Big(\Big(\frac{2}{3}, 1\Big]\Big) = \Xi^{-1}\Big(\Big(\frac{2}{3}, \infty\Big)\Big),$$
$$W_2 = \Xi^{-1}\Big(\Big(\frac{1}{3}, 1\Big]\Big) = \Xi^{-1}\Big(\Big(\frac{1}{3}, \infty\Big)\Big).$$

Множества W_1, W_2 — открытые, и

$$\Sigma^1 \subset W_1 \subset \operatorname{clos} W_1 \subset W_2 \subset \operatorname{clos} W_2 \subset U, \quad \zeta(W_1) = 1,\ \zeta(\mathbb{C} \setminus W_2) = 0 \tag{43}$$

(см. рис. 40). Произведение $\zeta(x)\varphi(x)$ непосредственно имеет смысл на $U \cap V$, но ввиду (43) можно считать, что оно определено и непрерывно на всей внешней области V, причём на $V \setminus W_2$ оно равно нулю. («Склеиваем» непрерывные функции

$$\begin{aligned} U \cap V \to \mathbb{R}, &\quad v \mapsto \zeta(v)\varphi(v), \\ V \setminus \operatorname{clos} W_2 \to \mathbb{R}, &\quad v \mapsto 0, \end{aligned}$$

определённые и непрерывные на двух открытых множествах $U \cap V$ и $V \setminus \operatorname{clos} W_2$, причём на их пересечении, равном $(U \cap V) \setminus \operatorname{clos} W_2$, они обе равны 0.) Положим

$$Q(x) = \frac{j(1) - x}{|j(1) - x|} e^{i\zeta(x)\varphi(x)}.$$

Q определено и непрерывно всюду на V. При этом $Q = P$ на $W_1 \cap V$ (ибо там $\zeta = 1$). Значит, положив

$$R(x) = \begin{cases} Q(x) & \text{при } x \in V, \\ P(x) & \text{при } x \in W_1, \end{cases}$$

мы корректно определим непрерывное отображение $V \cup W_1 \to \mathbb{S}^1$. (Как обычно, корректность определения следует из того, что там, где «об-

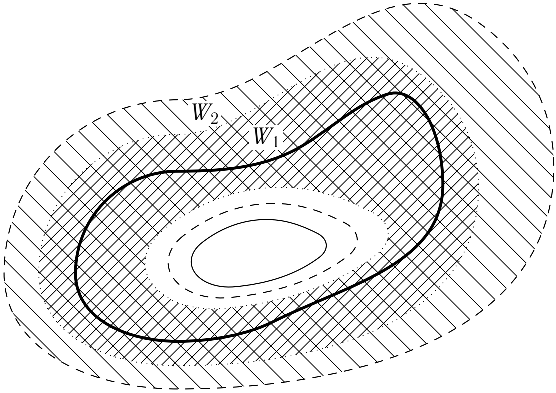


Рис. 40

ласти действия» обеих строк пересекаются, даваемые этими строками предписания совпадают, $Q(x) = P(x)$. Непрерывность следует из того, что множества, где определены и непрерывны «склеиваемые» отображения Q и P, являются открытыми.) Когда $|x|$ достаточно велико, то $R(x) = Q(x) = \dfrac{j(1) - x}{|j(1) - x|}$. Мы как раз и хотели построить непрерывное и нигде не равное нулю векторное поле на $V \cup \Sigma^1$ с такими свойствами (а построили его на $V \cup W_1 \supset V \cup \Sigma^1$).

Приложение

Специалистов, если они вообще будут держать в руках эту брошюру, может заинтересовать утверждение, что непрерывное отображение $F\colon A_1 \times A_2 \to \mathbb{S}^1$ не всегда гомотопно отображению вида $(z, w) \mapsto f(z)g(w)$. (Данное утверждение вполне может быть известным, но, видимо, общеизвестным его считать нельзя — я, во всяком случае, не встречал его в литературе.) Обращаясь к специалистам, я написал это приложение более кратко, чем основной текст; в частности, я не останавливался на геометрических пояснениях.

В приводимом ниже примере $A_1 = A_2$ является замыканием графика функции $y = \dfrac{1}{x} \sin \dfrac{1}{x}$, $x > 0$. Сам этот график обозначается через G, а присоединяемая к нему при замыкании ось ординат — через Y. На G вместо координат (x, y) удобно пользоваться координатой $s = \dfrac{1}{x}$, так что точка $z = (x, y) \in G$ имеет s-координату $s(z) = \dfrac{1}{x}$; в этих терминах $z = \left(\dfrac{1}{s}, s \sin s\right)$.

Положим при $0 \leqslant \xi \leqslant 1$

$$\alpha(\xi) = \begin{cases} 0 & \text{при } \xi \in \left[0,\ \frac{1}{3}\right], \\ 3x - 1 & \text{при } \xi \in \left[\frac{1}{3},\ \frac{2}{3}\right], \\ 1 & \text{при } \xi \in \left[\frac{2}{3},\ 1\right] \end{cases}$$

и при $0 \leqslant u,\ v \leqslant 2$

$$\beta(u,\ v) = \min(u,\ 1 - u,\ v,\ 1 - v), \qquad \gamma(u,\ v) = \alpha(\beta(u,\ v)).$$

Последняя функция определена и непрерывна в квадрате $0 \leqslant u,\ v \leqslant 2$, причём она равна нулю вне $\left[\frac{1}{3},\ \frac{5}{3}\right] \times \left[\frac{1}{3},\ \frac{5}{3}\right]$, а $\gamma(1,\ 1) = 1$. Продолжим $\gamma(u,\ v)$ на всю плоскость $\mathbb{R}^2$ как периодическую функцию от $u,\ v$ с периодом 2 по каждому аргументу. Так как первоначально эта функция была определена и непрерывна в квадрате $0 \leqslant u,\ v \leqslant 2$, а на его границе равнялась нулю, то при продолжении получится непрерывная функция на $\mathbb{R}^2$. Заметим, что при целых n

$$\gamma(n,\ n) = \begin{cases} \gamma(0,\ 0) = 0, & \text{если } n \text{ чётно}, \\ \gamma(1,\ 1) = 1, & \text{если } n \text{ нечётно}. \end{cases}$$

Наконец, положим для $z,\ w \in G$

$$\Psi(z,\ w) = \begin{cases} 2\pi\gamma\left(\frac{s(z)}{\pi},\ \frac{s(w)}{\pi}\right), & \text{если числа } s(z),\ s(w) \text{ оба лежат в одном} \\ & \text{и том же отрезке вида } [2\pi(n-1),\ 2\pi n] \\ & \text{с некоторым натуральным } n, \\ 0 & \text{в противном случае} \end{cases}$$

и для $z,\ w \in A_1$

$$F(z,\ w) = \begin{cases} 1, & \text{если } z \in Y \text{ (а } w \text{ любое) или } w \in Y \text{ (а } z \text{ любое)}, \\ e^{i\Psi(z,w)}, & \text{если } z,\ w \in G. \end{cases}$$

Отметим, что для $z_n = \left(\frac{1}{\pi n},\ 0\right)$ (n — натуральные числа; z_n являются точками из G)

$$\Psi(z_n,\ z_n) = 2\pi\gamma(n,\ n) = \begin{cases} 0, & \text{если } n \text{ чётно}, \\ 2\pi, & \text{если } n \text{ нечётно}. \end{cases} \tag{44}$$

Отображение $F\colon A_1 \times A_1 \to \mathbb{S}^1$ непрерывно. Действительно, непрерывность ограничений F на $G \times G$ и $(Y \times A_1) \cup (A_1 \times Y)$ очевидна. Остаётся убедиться, что если $z' \in Y$ и $w' \in A_1$ или $w' \in Y$ и $z' \in A_1$, а

$$(z_k,\, w_k) \to (z',\, w'), \quad z_k,\, w_k \in G,$$

то $F(z_k,\, w_k) \to 1$. Но если, скажем, $z' = (0,\, y) \in Y$ и $z_k \to z'$, $z_k \in G$, то

$$z_k = (x_k,\, y_k), \quad x_k \to 0, \;\; y_k \to y.$$
$$s(z_k) \to \infty, \quad s(z_k)\sin s(z_k) \to y, \quad \sin s(z_k) \to 0,$$

так что при больших k каждое $s(z_k)$ близко к некоторому $2\pi n_k$ с каким-то натуральным n_k. А между тем, когда $|s(z_k) - 2\pi n_k| < \frac{1}{3}$, то $\gamma(s(z_k),\, t) = 0$ при любом t. Поэтому при достаточно больших k

$$\Psi(z_k,\, w_k) = 0, \quad F(z_k,\, w_k) = 1.$$

Если бы F было гомотопно отображению вида $(z,\, w) \mapsto f(z)g(w)$, то отображение

$$(z,\, w) \mapsto \frac{F(z,\, w)}{f(z)g(w)}$$

было бы несущественным и у него имелась бы угловая функция $\Phi(z,\, w)$, так что F представлялось бы в виде

$$F(z,\, w) = f(z)g(w)e^{i\Phi(z,w)}.$$

В частности, при $z_0 = w_0 = (0,\, 0)$ (это точка из Y) мы имели бы при всех $z,\, w \in A_1$

$$\begin{aligned}
1 &= F(z,\, w_0) = f(z)g(w_0)e^{i\Phi(z,w_0)},\\
1 &= F(z_0,\, w) = f(z_0)g(w)e^{i\Phi(z_0,w)},\\
f(z) &= \frac{e^{-i\Phi(z,w_0)}}{g(w_0)}, \quad g(w) = \frac{e^{-i\Phi(z_0,w)}}{f(z_0)},\\
F(z,\, w) &= \frac{e^{i[-\Phi(z,w_0) - \Phi(z_0,w) + \Phi(z,w)]}}{f(z_0)g(w_0)}.
\end{aligned}$$

Меняя обозначения, можем написать, что $F(z,\, w) = e^{i\Phi(z,w)}$ с некоторой новой непрерывной функцией $\Phi\colon A_1 \times A_1 \to \mathbb{R}$. Но на связном множестве $G \times G$ мы имеем аналогичное представление для F с Ψ вместо Φ. Отсюда следует, что

$$\Phi(z,\, w) = \Psi(z,\, w) + 2\pi l$$

с некоторым постоянным натуральным l. Поэтому (см. (44))

$$\Phi(z_n,\, z_n) = \begin{cases} 2\pi l, & \text{если } n \text{ чётно},\\ 2\pi l + 2\pi, & \text{если } n \text{ нечётно}. \end{cases}$$

Это исключает возможность существования предела $\lim\limits_{n\to\infty}\Phi(z_n,\, z_n)$. Но ясно, что $z_n \to z_0 = (0,\, 0)$, и раз Φ непрерывна на всём $A_1 \times A_1$, то указанный предел должен существовать (и равняться $\Phi(z_0,\, z_0)$).

§ 9. Окончание доказательства теоремы Жордана

Итак, мы знаем, что $\mathbb{C} \setminus \Sigma^1$ несвязно и что каждая компонента связности характеризуется своим числом f_a. Дабы убедиться, что компонент не больше двух, достаточно показать, что $\mathrm{Ст}\, f_a$ может принимать только два значения. Сперва мы докажем, что возможны только значения -1, 0, 1, а потом — что значения -1 и 1 несовместимы, т. е. что если где-то $f_a = 1$, то нет точек b с $f_b = -1$.

Допустим, что $\mathrm{Ст}\, f_a = k$, $|k| > 1$. Пусть по-прежнему j — гомеоморфизм $\mathbb{S}^1 \to \Sigma^1$. Рассматривая j как петлю $\mathbb{S}^1 \to \mathbb{C}$, заметим, что в определении её порядка $\mathrm{Пор}(j,\, a)$ фигурирует отображение $z \mapsto \dfrac{j(z) - a}{|j(z) - a|} = f_a(j(z))$; поэтому $\mathrm{Пор}(j,\, a) = \mathrm{Ст}(f_a \circ j) = \mathrm{Ст}\, f_a = k$. Без ограничения общности можно считать, что $a = 0$.

Докажем, что если петля $f\colon \mathbb{S}^1 \to \mathbb{C}$ такова, что $\mathrm{Пор}(f,\, 0) = k$, $|k| > 1$, то у неё имеется самопересечение, т. е. существуют такие z_1, $z_2 \in \mathbb{S}^1$, что $z_1 \neq z_2$ и $f(z_1) = f(z_2)$.

Используя полярные координаты и комплексные числа, имеем $f(e^{i\varphi}) = r(\varphi)e^{i\psi(\varphi)}$, где $r(\varphi) = |f(e^{i\varphi})|$ — положительная периодическая функция с периодом 2π, а $\psi(\varphi)$ — угловая функция отображения $f/|f|\colon \mathbb{S}^1 \to \mathbb{S}^1$, так что она непрерывна и $\psi(\varphi + 2\pi) = \psi(\varphi) + 2\pi k$. Определим новое отображение $g\colon \mathbb{S}^1 \to \mathbb{C}$ формулой

$$g(e^{i\varphi}) = r(\varphi)e^{\frac{1}{k}i\psi(\varphi)}.$$

Поскольку представление точек $\mathbb{S}^1$ в виде $e^{i\varphi}$ неоднозначно, надо проверить, что это определение корректно, т. е. что при замене φ на $\varphi + 2\pi$ результат будет тот же. Но $r(\varphi)$ при этом не изменится, а

$$\frac{1}{k}\psi(\varphi + 2\pi) = \frac{1}{k}\psi(\varphi) + 2\pi,$$

так что $e^{\frac{1}{k}i\psi(\varphi)}$ тоже не изменится. Заодно мы видим, что $\mathrm{Пор}(g,\, 0) = 1 \neq 0$.

Повернём петлю g на угол $\dfrac{2\pi}{k}$ «в положительном направлении» (кавычки использованы потому, что при $k < 0$ имеется в виду поворот на угол $\dfrac{2\pi}{|k|}$ в отрицательном направлении), т. е. определим петлю $g_1(z) = f(z)e^{i\frac{2\pi}{k}}$. В § 5 мы показали, что $g(\mathbb{S}^1) \cap g_1(\mathbb{S}^1) \neq \varnothing$. Значит, имеются такие z_1, $z_2 \in \mathbb{S}^1$, что

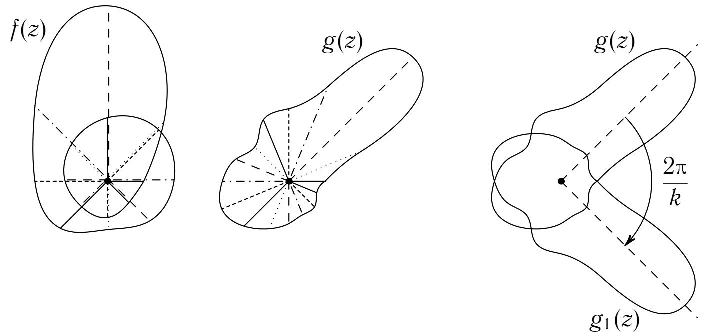


Рис. 41

$g(z_1) = e^{i\frac{2\pi}{k}} g(z_2)$ (откуда, кстати, видно, что $z_1 \neq z_2$). Это эквивалентно тому, что если φ_i суть аргументы z_i, то $r(\varphi_1) = r(\varphi_2)$ и

$$\frac{1}{k}\psi(\varphi_1) = \frac{2\pi}{k} + \frac{1}{k}\psi(\varphi_2) + 2\pi n \qquad \text{с некоторым целым } n.$$

Но тогда

$$\psi(\varphi_1) = \psi(\varphi_2) + 2\pi m \qquad \text{с некоторым целым } m,$$

откуда $f(z_1) = f(z_2)$. Тем самым доказано, что у петли f действительно имеется самопересечение.

Итак, для f_a остаются только три возможности: $\text{Ст} f_a = -1, 0, 1$. Во внешней области, конечно, $f_a = 0$. Может ли случиться, что $\text{Ст} f_a = 1$, $\text{Ст} f_b = -1$ для некоторых a, b?

Пусть это так. Используем вспомогательный гомеоморфизм

$$h\colon \mathbb{C} \setminus \{b\} \to \mathbb{C} \setminus \{1\}, \qquad h(w) = \frac{w-a}{w-b}.$$

(Это действительно гомеоморфизм — почему?) При нём Σ^1 — образ $\mathbb{S}^1$ при гомеоморфизме j — переходит в образ Σ' стандартной окружности $\mathbb{S}^1$ при гомеоморфизме $j' = h \circ j$. Это снова топологическая окружность. Для неё роль f_a (с $a = 0$) играет отображение

$$f_0'\colon \Sigma' \to \mathbb{S}^1, \qquad \zeta \mapsto \frac{\zeta}{|\zeta|}.$$

Если $\zeta = h(w)$, $w \in \Sigma^1$, то $f_0'(\zeta) = \dfrac{\zeta}{|\zeta|} = \dfrac{w-a}{w-b} : \dfrac{|w-a|}{|w-b|} = \dfrac{f_a(w)}{f_b(w)}$. Поэтому

$$\text{Ст} f_0' = \text{Ст} f_a - \text{Ст} f_b = 2.$$

Но мы видели, что это невозможно.

Литература

[1] *Чернавский А. В.* Теорема Жордана // Математическое просвещение. Третья сер. Вып. 3. С. 142—157.

[2] *Eilenberg S.* Transformations continues en circonference et la topologie du plan // Fund. Math. 1936. V. 26. P. 61—112.

[3] *Дьедонне Ж.* Основы современного анализа. М.: Мир, 1964.

[4] *Стинрод Н.*, *Чинн У.* Первые понятия топологии. М.: Мир, 1967.

[5] *Курант Р.*, *Роббинс Г.* Что такое математика? М.: МЦНМО, 2001.

[6] *Болтянский В. Г.*, *Ефремович В. А.* Наглядная топология. М.: Наука, 1982.

[7] *Виленкин Н. Я.* Рассказы о множествах. М.: МЦНМО, 2003.

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

[9] *Шафаревич И. Р.* О решении уравнений высших степеней (метод Штурма). М.: ГИТТЛ, 1954.

Оглавление