В. В. Прасолов, Ю. П. Соловьев

Эллиптические функции *и* алгебраические уравнения

Москва „Факториал“ 1997

ББК 22.132
П 70
УДК 511.3

П 70 **Прасолов В. В., Соловьев Ю. П.** Эллиптические функции и алгебраические уравнения. — М.: Изд-во «Факториал», 1997. — 288 с. — ISBN 5-88688-018-6.

Книга представляет собой вводный курс в теорию эллиптических функций и эллиптических кривых и предназначена для первого знакомства с предметом.

Основные вопросы, рассматриваемые в книге — это геометрия кубических кривых, эллиптические функции и их свойства, эллиптические интегралы, теоремы сложения эллиптических функций и интегралов, теорема Абеля о лемнискате, теорема Морделла, тэта-функции, кривые Серре.

Кроме того, впервые в учебной литературе, приводится вывод теоремы Ферма из некоторых гипотез об эллиптических кривых.

В книге подробно изложена классическая теория решения общего алгебраического уравнения пятой степени в тэта-функциях.

Ил. 46. Библиогр. 80.

РФФИ

Издание осуществлено при финансовой поддержке Российского фонда фундаментальных исследований. Проект № 96-01-14036.

Научное издание

Прасолов Виктор Васильевич
Соловьев Юрий Петрович

Эллиптические функции и алгебраические уравнения.

Редактор *Попеленский Ф. Ю.*
Художник *Рожковская Н. А.*

Формат 60 × 90/16. Гарнитура литературная. Усл. печ. л. 18. Бумага офсетная № 1. Подписано к печати 11.6.1997. Тираж 1300 экз. Заказ № 1790

Издательство «Факториал», 17449, Москва, а/я 331; ЛР № 063537 от 22.07.1994.

Оригинал-макет подготовлен с использованием макропакета **AP**-TEX.

Отпечатано во 2-й типографии издательства «Наука». 121099, Москва Г-99, Шубинский пер., 6.

ISBN 5-88688-018-6

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ

В июне 1796 г., в «Литературной газете», издававшейся в Иене, была помещена следующая заметка.

НОВЫЕ ОТКРЫТИЯ

Всякому начинающему геометру известно, что можно геометрически, т. е. циркулем и линейкой, строить разные правильные многоугольники, а именно треугольник, пятиугольник, пятнадцатиугольник и те, которые получаются из каждого из этих путем последовательного удвоения числа его сторон. Это было уже известно во времена Евклида, и, как кажется, с тех пор господствовало убеждение, что область элементарной геометрии дальше не распространяется: по крайней мере, я не знаю удачной попытки распространить ее в эту сторону. Тем более кажется мне заслуживающим внимания открытие, что кроме этих правильных многоугольников может быть геометрически построено еще множество других, например, семнадцатиугольник. Это открытие является собственно лишь следствием одной еще не совсем законченной большой теории. Как только она обретет завершенность, то будет предложена публике.

К. Ф. Гаусс из Брауншвейга,
студент-математик в Гёттингене.

Законченный вид эта теория обрела через пять лет. Она была опубликована в седьмом разделе знаменитых «Disquisitiones Arithmeticae» («Арифметических исследований») Гаусса, увидевших свет в 1801 г. Гаусс доказал, что если число n сторон правильного многоугольника имеет вид $n = 2^a p_1 \dots p_k$, где p_i — различные простые числа Ферма, т. е. простые числа вида $2^{2^m}+1$, то многоугольник может быть построен циркулем и линейкой. На алгебраическом языке это утверждение означает, что для числа n указанного вида уравнение $x^n - 1 = 0$ может быть решено в квадратных радикалах. Доказательство теоремы Гаусса основано на изящной алгебраической теории, которая послужила краеугольным камнем для теории Галуа, созданной тридцать лет спустя после появления «Арифметических исследований».

В упомянутом седьмом разделе «Арифметических исследований» кроме теории деления круга, т. е. алгебраической теории круговых функций, содержится также краткое замечание Гаусса о том, что разработанный им метод может быть применен и к некоторым высшим трансцендентным функциям, в частности, к функциям, связанным с интегралами вида $\int \frac{dx}{\sqrt{1-x^4}}$. Это замечание было отправной точкой для исследований Абеля, который в 1827 г. доказал, что на те же самые гауссовские n частей можно разделить с помощью циркуля и линейки лемнискату Бернулли. Для этого Абелю пришлось усовершенствовать метод Гаусса, и главное, создать новую математическую дисциплину — теорию эллиптических функций. Эта теория и ее геометрический двойник — теория эллиптических кривых — заняли одно из центральных мест в математике, синтезировав самые различные ее ветви. Несмотря на свой почтенный возраст, теория эллиптических функций и эллиптических кривых остается живой и бурно развивающейся областью математики и служит неиссякаемым источником исследовательской техники, задач и гипотез. Достаточно сказать, что в последнее десятилетие эллиптические функции и кривые стали предметом пристального внимания специалистов даже в таких неклассических областях, как алгебраическая топология и квантовая теория поля, а совсем недавно с помощью теории эллиптических кривых получено доказательство последней теоремы Ферма.

Предмет настоящей книги — это теория эллиптических интегралов, эллиптических функций и эллиптических кривых, развитая в той степени, которая необходима для доказательства теоремы Абеля о лемнискате. Основные вопросы, рассматриваемые в книге — это геометрия кубических кривых, эллиптические функции и их свойства, эллиптические интегралы, теоремы сложения эллиптических функций и интегралов, дуги алгебраических кривых, выражаемые в эллиптических интегралах, теорема Абеля о лемнискате. Кроме того, мы рассматриваем арифметические свойства эллиптических кривых, теорему Морделла, вывод теоремы Ферма из некоторых гипотез об эллиптических кривых, тэта-функции и решение общего алгебраического уравнения пятой степени в тэта-функциях. Иными словами, книга представляет собой вводный курс в теорию эллиптических функций и эллиптических кривых и предназначена для первого знакомства с предметом. В то же время, мы надеемся, что она окажется интересной и для профессионалов.

Основу книги составили три лекции, прочитанные одним из нас (Ю. С.) в 1991/92 учебном году в цикле студенческих чтений Московского математического общества. Изложенный в книге материал был собран воедино для специального курса, реализованного вторым автором (В. П.) в 1992/93 учебном году в Московском независимом университете. Во время работы над книгой В. прасолов получал финансовую поддержку от Российского Фонда Фундаментальных Исследований (проект № 95–01–00846). Предварительная версия книги была опубликована в виде лекционных заметок *) . Мы признательны редактору книги Ф. Ю. Попеленскому за плодотворное сотрудничество.

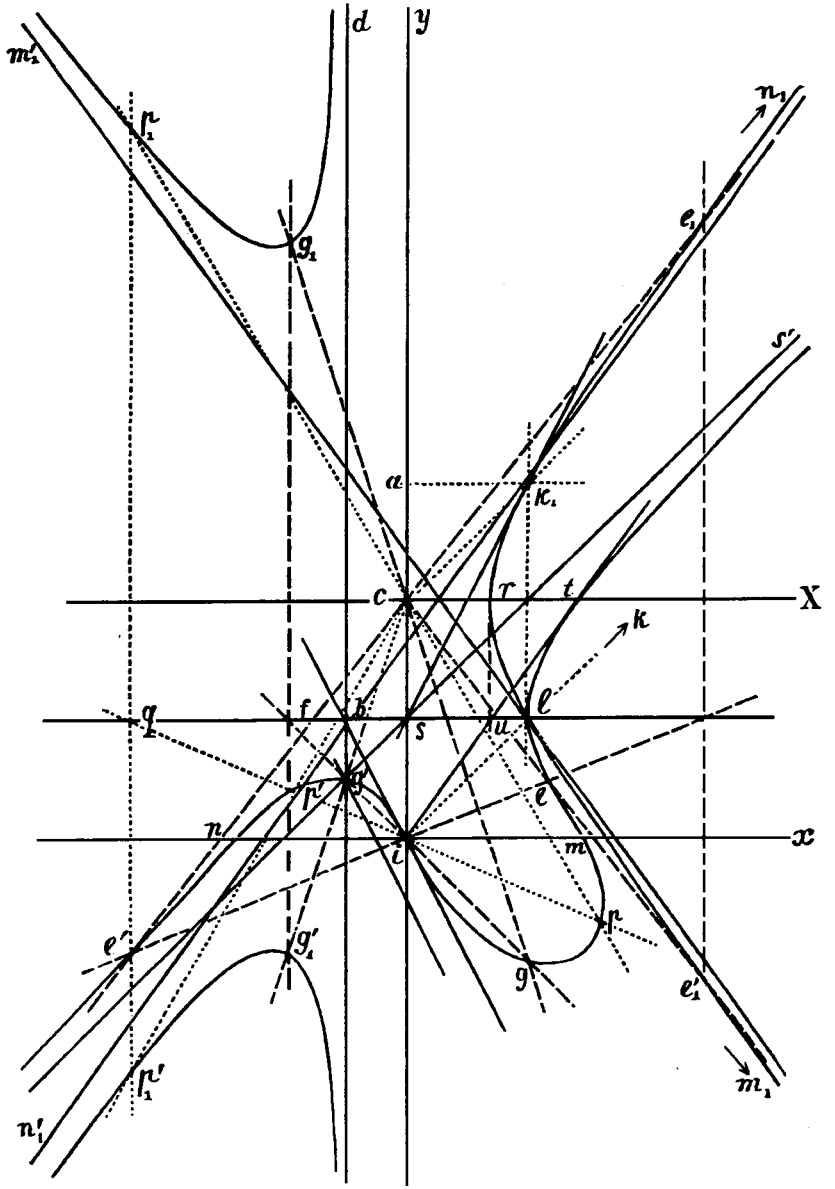
При написании книги мы использовали обширную литературу — как классические труды, так и различные современные работы. Большое влияние на нас оказала замечательная статья М. Роузена из «American Mathematical Monthly» [В14], содержащая современное доказательство теоремы Абеля. Много полезного мы почерпнули из прекрасных руководств Вебера [Б11], Хьюзмоллера [Б31], Коблица [Б17] и Степанова [Б27].

Книга не предполагает у читателя никаких предварительных знаний, выходящих за пределы младших курсов математических специальностей университетов, и ориентирована на самый широкий круг читателей: студентов-математиков и физиков, школьных учителей и даже школьников старших классов. Надеемся, что читатель, познакомившийся с предметом настоящей книги, сможет ощутить то очарование тонкого искусства, которое испытывали мы, разбирая работы старых мастеров.

г. Москва, октябрь 1996 г.

*) Прасолов В. В., Соловьев Ю. П. Эллиптические функции. Специальный курс. — М.: Издательство МК НМУ, 1993.

Прасолов В. В., Соловьев Ю. П. Алгебраические уравнения и тэта-функции. — М.: Издательство МК НМУ, 1994.



ГЛАВА 1

ГЕОМЕТРИЯ КУБИЧЕСКИХ КРИВЫХ

§ 1. Сложение точек кубической кривой

Плоской алгебраической кривой называется множество точек $(x, y) \in \mathbb{R}^2$, удовлетворяющих уравнению $f(x, y) = 0$, где $f(x, y)$ — ненулевой многочлен.

На некоторых плоских кривых существуют естественные групповые законы сложения точек. Простейшими примерами таких кривых являются прямая и окружность.

Для того, чтобы определить сложение точек на прямой, зафиксируем на ней некоторую точку O. Суммой точек прямой X и Y будем считать такую точку Z, что $\overrightarrow{OZ} = \overrightarrow{OX} + \overrightarrow{OY}$. Очевидно, что точки прямой образуют коммутативную группу относительно этой операции сложения.

Суммой двух точек $(r\cos\alpha, r\sin\alpha)$ и $(r\cos\beta, r\sin\beta)$, окружности $x^2 + y^2 = r^2$ будем считать точку $(r\cos(\alpha+\beta), r\sin(\alpha+\beta))$.

Этот закон сложения точек геометрически можно проинтерпретировать следующим образом. Пусть $E = (r, 0)$, A и B — произвольные точки единичной окружности. Проведем через точку E прямую, параллельную прямой AB; она пересекает окружность в точке C. Будем считать точку C суммой точек A и B (рис. 1).

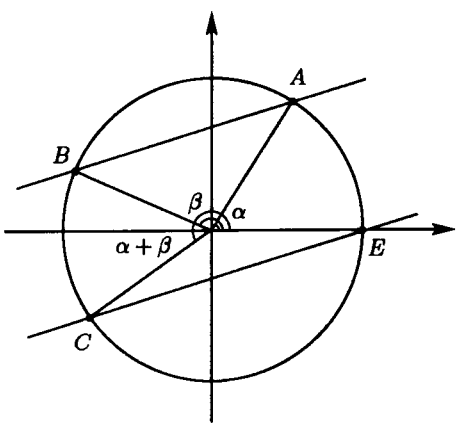


Рис. 1

В таком виде определение годится для любой коники, т. е. кривой второго порядка.

Фиксируем на конике некоторую точку E и будем считать суммой точек A и B точку, в которой прямая, проведенная через точку E параллельно прямой AB, вторично пересекает конику. Коммутативность полученной операции очевидна, нулевым элементом служит точка E. Для нахождения элемента $-A$ нужно провести через точку A прямую, параллельную касательной в точке E. Закон ассоциативности проверяется несколько сложнее. Рассмотрим на конике точки A, B и C. Обозначим точки $A+B$ и $B+C$ через P и Q соответственно. Равенство

$$A+(B+C)=(A+B)+C$$

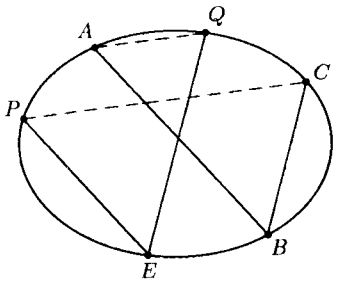


Рис. 2

эквивалентно следующему утверждению: «Если A, B, C, E, P и Q — точки коники, причем $AB \parallel EP$ и $BC \parallel EQ$, то $AQ \parallel CP$». Это — частный случай теоремы Паскаля о шестиугольнике, вписанном в конику.

Для параболы $y=x^2$ с фиксированной точкой $E=(0,0)$ суммой точек (x_1,y_1) и (x_2,y_2) будет точка $(x_1+x_2, y_1+y_2+2x_1x_2)$. Для гиперболы $x^2-y^2=1$ с фиксированной точкой $E=(1,0)$ суммой точек (x_1,y_1) и (x_2,y_2) будет точка $(x_1x_2+y_1y_2, y_1x_2+y_2x_1)$. При параметризации $x=\operatorname{ch} t$, $y=\operatorname{sh} t$ это сложение соответствует сложению параметра t.

Кубической кривой называется плоская алгебраическая кривая $\sum\limits_{i,j} a_i x^i y^j = 0$, где наибольшее значение $i+j$ равно трем. На любой неособой кубической кривой (свойства этих кривых мы подробно обсудим в § 3) тоже существует естественный закон сложения точек.

Закон сложения несовпадающих точек произвольной кубической кривой можно определить следующим образом. Отметим на ней произвольную точку E. Чтобы сложить точки A и B, проведем прямую AB. Она пересечет кубическую кривую в некоторой точке X. Точку пересечения прямой XE с кубической кривой будем считать суммой точек A и B (рис. 3).

В определении сложения дважды использовано следующее свойство кубической кривой: если прямая пересекает ее в двух точках, то она пересекает ее еще ровно в одной точке. Это свойство кажется почти очевидным. В самом деле, из уравнения прямой $ax+by+c=0$ можно выразить x или y и подставить в уравнение кубической кривой. Получим уравнение третьей степени. По условию у него есть два вещественных корня, а значит, должен быть и третий вещественный корень.

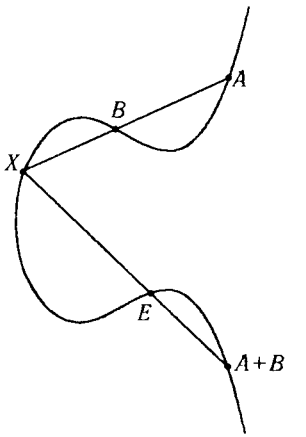


Рис. 3

В действительности все не так просто. И дело не только в том, что уравнение может иметь кратные корни. Его степень может оказаться меньше трех. С этим нам еще придется разобраться (см. § 2), потому что иначе операция сложения точек будет неполноценной: складывать можно будет не все точки.

Коммутативность полученной операции очевидна. Легко проверить также, что E — нулевой элемент. Докажем ассоциативность операции.

Равенство

$$(A+B)+C=(A+C)+B$$

эквивалентно тому, что точка пересечения прямых, соединяющих точки $A+B$ и C, $A+C$ и B, лежит на кубической кривой (рис. 4). Обозначим изображенные на рис. 4 прямые следующим образом:

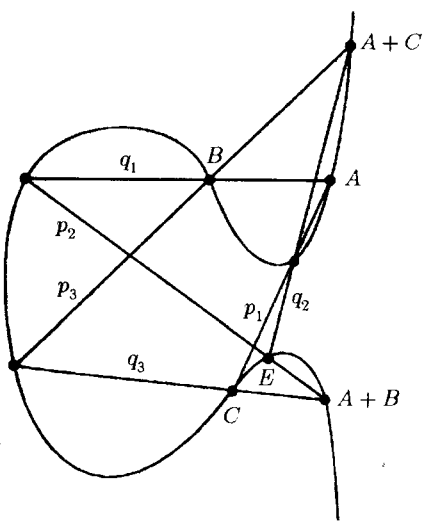


Рис. 4

$$p_1=AC,\quad p_2=E(A+B),$$
$$p_3=B(A+C),$$
$$q_1=AB,\quad q_2=E(A+C),$$
$$q_3=C(A+B).$$

Будем считать, что все точки пересечения прямых p_i и q_j попарно

различны. Тогда утверждение, которое нужно доказать, можно сформулировать следующим образом:

Теорема 1. *Пусть A_{ij} — точка пересечения прямых p_i и q_j, где $1 \leqslant i, j \leqslant 3$, причем точки A_{ij} попарно различны. Про все точки A_{ij}, кроме точки A_{33}, известно, что они лежат на некоторой кубической кривой. Тогда точка A_{33} тоже лежит на этой кубической кривой.*

Доказательство. Пусть $p_i(x,y) = 0$ и $q_j(x,y) = 0$ — уравнения прямых p_i и q_j. Тогда уравнение третьей степени $p_1p_2p_3 = 0$ задает тройку прямых p_1, p_2 и p_3, а уравнение $q_1q_2q_3 = 0$ задает тройку прямых q_1, q_2 и q_3. Кубическая кривая $\alpha p_1p_2p_3 + \beta q_1q_2q_3 = 0$ проходит через все точки A_{ij}. Оказывается, что в таком виде можно представить уравнение любой кубической кривой, проходящей через восемь из девяти точек A_{ij}. Докажем это. Выберем в качестве осей координат прямые p_1 и q_1, т. е. будем считать, что $p_1(x,y) = y$ и $q_1(x,y) = x$. Пусть рассматриваемая кубическая кривая задается уравнением $P(x,y) = 0$. Функции $P(0,y)$ и $yp_2(0,y)\,p_3(0,y)$ обращаются в нуль в трех точках A_{11}, A_{21} и A_{31}, лежащих на оси Oy (рис. 5). Кроме того, эти функции — многочлены степени не более трех. Следовательно,

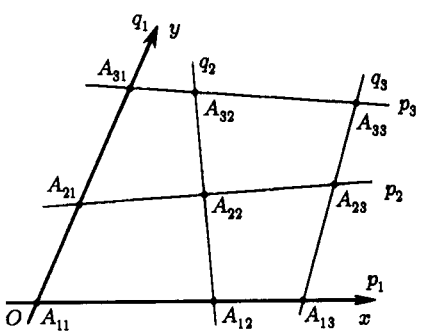


Рис. 5

$$P(0,y) = \alpha y\, p_2(0,y)\, p_3(0,y).$$

Аналогично

$$P(x,0) = \beta x\, q_2(x,0)\, q_3(x,0).$$

Рассмотрим многочлен

$$Q(x,y) = P(x,y) - \alpha y\, p_2(x,y)\, p_3(x,y) - \beta x\, q_2(x,y)\, q_3(x,y).$$

Ясно, что

$$Q(0,y) = P(0,y) - \alpha y\, p_2(0,y)\, p_3(0,y) = 0.$$

Многочлен

$$a_0(y) + a_1(y)\,x + a_2(y)\,x^2 + \ldots$$

тождественно равен нулю при $x = 0$ лишь в том случае, когда $a_0(y) = 0$, т. е. этот многочлен делится на x. Аналогичные рассуждения показывают, что многочлен $Q(x,y)$ делится и на y, т. е.

$$Q(x,y) = xy\,Q_1(x,y).$$

Степень многочлена $Q(x,y)$ не превосходит трех, поэтому $Q_1(x,y)$ — линейная функция или константа. Вспомним теперь, что в точках A_{22}, A_{23} и A_{32} обращаются в нуль многочлены P, p_2p_3 и q_2q_3, а значит, в этих точках обращается в нуль и многочлен Q. А так как во всех этих точках $xy \neq 0$, то в них обращается в нуль и линейная функция Q_1. Точки A_{22}, A_{23} и A_{32} не лежат на одной прямой, а для ненулевой линейной функции f уравнение $f(x,y) = 0$ определяет прямую. Следовательно, $Q_1 = 0$, т. е. $P = \alpha p_1p_2p_3 + \beta q_1q_2q_3$. В частности, точка A_{33} лежит на кривой $P(x,y) = 0$. □

Мы доказали также, что любая кубическая кривая, проходящая через точки A_{ij}, задается уравнением $\alpha p_1p_2p_3 + \beta q_1q_2q_3 = 0$. Иными словами, такие кривые образуют однопараметрическое семейство.

С помощью теоремы 1 можно получить очень простые доказательства двух классических теорем.

Теорема 2 (Паскаль). *Точки пересечения противоположных сторон вписанного шестиугольника лежат на одной прямой.*

Доказательство. Пусть $p_1 = AB$, $q_1 = BC$, $p_2 = EF$, $q_2 = DE$, $p_3 = CD$, $q_3 = AF$ (рис. 6). В качестве кубической кри-

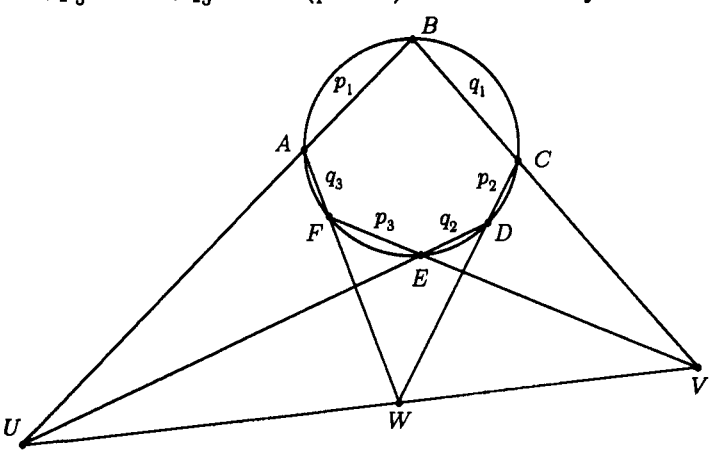


Рис. 6

вой возьмем кривую, заданную уравнением $Ql = 0$, где $Q = 0$ — уравнение окружности, $l = 0$ — уравнение прямой UV, а U и V — точки пересечения прямых p_1 и q_2, p_2 и q_1 соответственно. Пусть W — точка пересечения прямых p_3 и q_3. Про все остальные точки пересечения прямых p_i и q_j известно, что они лежат на кривой $Ql = 0$. Следовательно, точка W тоже лежит на этой кривой, причем именно на прямой l, а не на окружности. □

Вместо окружности $Q = 0$ можно взять любую кривую второго порядка. В частности, можно считать, что $Q = pq$, где p и q — линейные функции. В этом случае мы получим теорему Паппа.

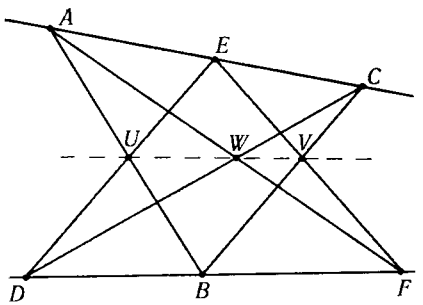


Рис. 7

Теорема 3 (Папп). *На прямой p взяты точки A, C и E, на прямой q взяты точки B, D и F. Прямые AB и DE, BC и EF, AF и CD пересекаются в точках U, V и W соответственно (рис. 7). Тогда точки U, V и W лежат на одной прямой.*

Доказательство. Дословно повторим доказательство теоремы Паскаля, считая, что $Q = pq$ — уравнение, задающее пару прямых p и q. □

Теорему 1 иногда приходится применять и в тех случаях, когда некоторые из точек A_{ij} совпадают. Поэтому следует понять, как нужно изменить ее формулировку, чтобы она оставалась верной и в таких ситуациях. Попарное различие точек A_{ij} использовалось для обоснования двух утверждений:

1) в точках A_{22}, A_{23} и A_{32} величина xy отлична от нуля, поэтому линейная функция Q_1 в них обращается в нуль;

2) эти точки не лежат на одной прямой, поэтому $Q_1 \equiv 0$.

В процессе доказательства теоремы 1 используются лишь ограничения многочлена P на прямые p_i и q_j. Поэтому можно ожидать, что вместо попарного различия точек A_{ij} достаточно предположить следующее: «Если две, соответственно три, точки A_{ij}, лежащие на прямой p_i или q_j совпадают, то ограничение многочлена P на эту прямую имеет в этой точке корень кратности два, соответственно три». Это изменение относится и к точке A_{33}.

Покажем, что так сформулированная теорема 1 остается верной. Доказательство того, что многочлен

$$Q = P - \alpha p_1 p_2 p_3 - \beta q_1 q_2 q_3$$

делится на $xy = p_1 q_1$, сохраняется без изменений. Если $A_{ij} = A_{ik} = A$, то в точке A ограничение P на p_i имеет корень кратности два, ограничение $p_1 p_2 p_3$ тождественно равно нулю, а ограничение $q_1 q_2 q_3$ имеет корень кратности два, так как $q_j(A_{ij}) = 0$ и $q_k(A_{ik}) = 0$. Следовательно, в точке A ограничение Q на p_i имеет корень кратности два. Для прямой q_j, а также в случае трех совпадающих точек рассуждения аналогичны. Теперь ясно, что в точках A_{22}, A_{23} и A_{32} линейная функция Q_1 по-прежнему обращается в нуль. В том случае, когда некоторые из этих точек совпадают, нужно воспользоваться тем, что ненулевая линейная функция на прямой не может иметь корень кратности два. Что же касается утверждения теоремы 1, то ясно, что для ограничения функции $\alpha p_1 p_2 p_3 + \beta q_1 q_2 q_3$ на прямую p_3 кратность корня в точке A_{33} равна количеству прямых q_j, проходящих через эту точку.

Для прямой l, касающейся кривой $F(X) = 0$ в точке X_0, ограничение F на l имеет в точке X_0 корень кратности два. В самом деле, пусть точка X_1 движется по этой кривой к точке X_0. Ограничение функции F на прямую X_0X_1 имеет корни X_0 и X_1. В предельном положении прямая X_0X_1 совпадает с l и корни X_0 и X_1 сливаются в один корень кратности два (рис. 8, а). Слияние трех корней происходит на касательной в точке перегиба (рис. 8, б). Более подробно точки кратного пересечения прямой и кубической кривой мы обсудим в § 3.

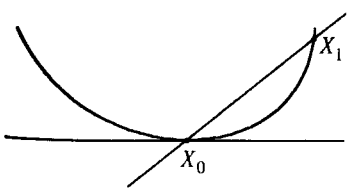


Рис. 8 *а)*

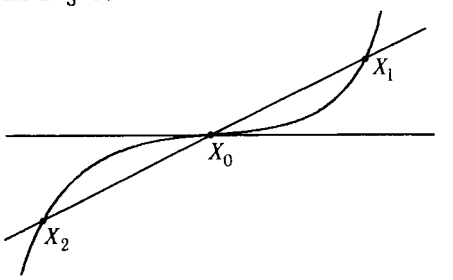


Рис. 8 *б)*

Для кривых степени $n \geqslant 3$ теорему 1 можно обобщить следующим образом.

Теорема 4. *Пусть A_{ij} — точка пересечения прямых p_i и q_j, где $1 \leqslant i, j \leqslant n$, причем точки A_{ij} попарно различны. Про все точки A_{ij}, для которых $i+j < n+3$, известно, что они лежат на кривой степени n. Тогда и остальные точки A_{ij} тоже лежат на этой кривой.*

Доказательство. Проведем индукцию по степени кривой n. База индукции $n = 3$ была рассмотрена при доказательстве теоремы 1. Пусть теперь $n > 3$. Выберем в качестве осей координат прямые p_1 и q_1. Пусть кривая, о которой идет речь в условии теоремы, задается уравнением $P_n(x, y) = 0$. Тогда $P_n(0, y) = \alpha p_1 \dots p_n$ и $P_n(x, 0) = \beta q_1 \dots q_n$. Рассмотрим многочлен $Q_n = P_n - \alpha p_1 \dots p_n - \beta q_1 \dots q_n$. Достаточно доказать, что $Q_n \equiv 0$. Легко проверить, что Q_n делится на $xy = p_1 q_1$, т. е. $Q_n = p_1 q_1 Q_{n-2}$. Остается доказать, что ненулевой многочлен Q_{n-2} степени не более $n-2$ не может обращаться в нуль в точках A_{ij}, где $i, j \geqslant 2$ и $i+j < n+3$ (на рис. 9 они изображены для $n = 5$). Предположим, что такой ненулевой многочлен Q_{n-2} существует. Его ограничение на прямую p_2 обращается в нуль в $n-1$ точке A_{22}, A_{23}, ..., A_{2n}. Следовательно, ограничение Q_{n-2} на эту прямую тождественно равно нулю, т. е. $Q_{n-2} = p_2 Q_{n-3}$. Многочлен Q_{n-3} обращается в нуль в точках, образующих аналогичную конфигурацию меньшего размера (на рис. 9 эти точки обведены). Эти рассуждения показывают, как делается шаг индукции. □

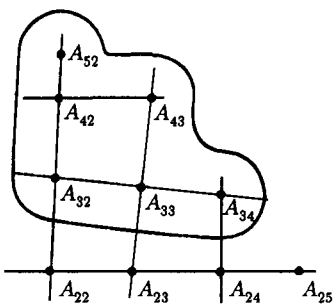


Рис. 9

Метод доказательства геометрических теорем с помощью семейства кривых $p_1 p_2 p_3 + \mu q_1 q_2 q_3 = 0$ был разработан немецким математиком Юлиусом Плюккером (1801–1868). Идея представить тройку прямых как вырожденную кубическую кривую оказалась весьма плодотворной. Доказательства многих сложных геометрических теорем сводились теперь к умелому подбору коэффициента μ, который стал часто появляться в статьях Плюккера. Такая алгебраизация геометрии не всем пришлась по душе. Якоб Штейнер (1796–1863), один из крупнейших геометров того времени, отказывался даже приписывать знаки геометрическим величинам и предпочитал вместо этого рассматривать различные варианты расположения точек. И при этом ему порой удавалось получать более тонкие и глубокие результаты, чем Плюккеру. О новых алгебраических методах в геометрии Штейнер отзывался неодобрительно.

Задачи.

1. Прямые AB и CD пересекаются в точке P, BC и AD — в точке Q. Кубическая кривая проходит через точки A, B, C, D, P, Q. Докажите, что касательные в точках P и Q пересекаются в точке, лежащей на этой кривой.

Указание. Примените теорему 1 в случае, когда $A_{31} = A_{32}$ и $A_{13} = A_{23}$.

2. Прямая пересекает кубическую кривую в точках A, B и C. Касательные в точках A, B и C пересекают кривую в точках A_1, B_1 и C_1. Докажите, что точки A_1, B_1 и C_1 лежат на одной прямой.

Указание. Примените теорему 1 в случае, когда $p_1 = p_2$.

3. Восьмиугольник со сторонами $l_1, \dots, l_8$ вписан в конику. Докажите, что восемь точек пересечения прямых l_i и l_j, где $j - i \equiv 3 \pmod 8$, лежат на одной конике.

Указание. Пусть $p_i = l_{2i-1}$, $q_i = l_{2i}$, $C_1 = 0$ — исходная коника, $C_2 = 0$ — коника, проходящая через пять из восьми оставшихся точек пересечения прямых p_i и q_j. Примените теорему 2 к кривой $C_1 C_2 = 0$.

4. Докажите,что уравнение любой кривой степени $n-1$, проходящей через все точки пересечения прямых $p_1 = 0, \dots, p_n = 0$, имеет вид

$$p_1 \dots p_n \left(\frac{\lambda_1}{p_1} + \dots + \frac{\lambda_n}{p_n} \right) = 0,$$

где $\lambda_1, \dots, \lambda_n$ — некоторые константы. (Предполагается, что все точки пересечения попарно различны.)

Указание. Пусть $C = 0$ — уравнение такой кривой. Рассмотрим прямую l, не проходящую через точки пересечения прямых p_i. Числа можно подобрать так, что во всех n точках пересечения прямой l с прямыми p_i выполняется равенство

$$C - p_1 \dots p_n \left(\frac{\lambda_1}{p_1} + \dots + \frac{\lambda_n}{p_n} \right) = 0.$$

Такое же равенство выполняется тогда в n точках любой из прямых p_i.

§ 2. Прямые и кривые на проективной плоскости

В предыдущем параграфе упоминалось о том, что сложение точек кубической кривой определено, вообще говоря, не для всех точек. Покажем это на примере кривой

$$y^2 = x(x-1)(x-2) \tag{2.1}$$

(рис. 10). Подставив в (2.1) уравнение прямой $x = 1/2$, получим $y^2 = 3/8$. Степень этого уравнения равна двум, а не трем. Значит, прямая $x = 1/2$ пересекает кривую (2.1) лишь в двух точках, причем пересечения не кратные. Попытка сложить эти точки не увенчается успехом. Если (x, y) — точка кривой (2.1), то

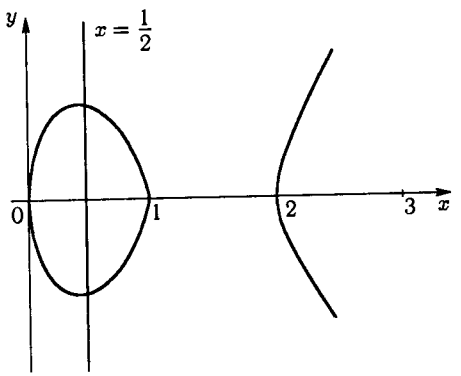


Рис. 10

$$\lim_{x\to\infty} \frac{x}{y} = \lim_{x\to\infty} \frac{x}{\sqrt{x^3}} = \lim_{x\to\infty} \frac{1}{\sqrt{x}} = 0.$$

Естественно предположить, что кривая (2.1) и прямая $x = 1/2$ проходят еще через одну точку — «бесконечно удаленную» точку в направлении оси Oy. Более того, недостающие точки пересечения могут оказаться на бесконечно удаленной прямой. Попытаемся к точкам обычной плоскости для каждого направления l добавить точку, в которой пересекаются параллельные прямые направления l. Сделать это нужно еще и потому, что иначе наши формулировки и доказательства теорем Паппа и Паскаля будут не совсем аккуратными. Мы ведь всегда предполагали,что рассматриваемые прямые пересекаются, но они могут быть и параллельными. Можно, конечно, отдельно рассматривать случаи, когда некоторые прямые пересекаются, а некоторые параллельны. Но это весьма утомительно, потому что

каждый случай требует не только отдельной формулировки, но и отдельного доказательства.

Рассмотрим некоторую точку O, не лежащую в плоскости П, и сопоставим каждой точке $A \in$ П прямую OA. Прямой $l \in$ П будет при этом сопоставлена вся плоскость Ol, содержащая точку O и прямую l, за исключением прямой l', проходящей через точку O параллельно l. Если прямая $l_1 \in$ П параллельна прямой l, то плоскости Ol и Ol_1 пересекаются по прямой l'. Ясно также, что если точка A удаляется по прямой l в бесконечность, то предельным положением прямой OA будет прямая l'.

Определим *проективную плоскость* $\mathbb{R}\mathrm{P}^2$ следующим образом. Назовем точками $\mathbb{R}\mathrm{P}^2$ прямые, проходящие через точку O. Прямыми $\mathbb{R}\mathrm{P}^2$ назовем плоскости, проходящие через точку O. При этом прямые, параллельные плоскости П, соответствуют бесконечно удаленным точкам плоскости П, а плоскость, параллельная плоскости П, соответствует бесконечно удаленной прямой плоскости П. На проективной плоскости любые две прямые пересекаются в одной точке. Проективные прямые, соответствующие параллельным прямым плоскости П, пересекаются в точке бесконечно удаленной прямой. Если про плоскость П забыть, то бесконечно удаленные точки ничем не будут отличаться от остальных точек.

Для работы с алгебраическими кривыми нужно ввести координаты на проективной плоскости. Будем считать, что точка O расположена в начале системы координат, а плоскость П задана уравнением $z = 1$. Прямая, проходящая через точку O, состоит из точек вида $(\lambda x, \lambda y, \lambda z)$, где числа x, y, z фиксированы. Поэтому точками $\mathbb{R}\mathrm{P}^2$ можно назвать ненулевые тройки вещественных чисел (x, y, z), причем тройки (x, y, z) и $(\lambda x, \lambda y, \lambda z)$ считаются эквивалентными. Бесконечно удаленная прямая задается уравнением $z = 0$.

В определении проективной плоскости числа x, y, z и λ можно считать комплексными. В этом случае получим определение *комплексной проективной плоскости* $\mathbb{C}\mathrm{P}^2$. Геометрия алгебраических кривых в $\mathbb{C}\mathrm{P}^2$ существенно проще, чем в $\mathbb{R}\mathrm{P}^2$. Это связано с тем, что над полем $\mathbb{C}$ любой многочлен n-й степени имеет с учетом кратности ровно n корней.

Кривой

$$y^2 = x(x-1)(x-2)$$

можно сопоставить кривую

$$y^2z = x(x-z)(x-2z)$$

на проективной плоскости. В самом деле, это уравнение действительно задает кривую на проективной плоскости, так как точки (x,y,z) и $(\lambda x,\lambda y,\lambda z)$ одновременно либо удовлетворяют ему, либо не удовлетворяют. Кроме того, на плоскости П, заданной уравнением $z=1$, оба уравнения совпадают. Аналогичным образом алгебраической кривой $\sum a_{ij}x^iy^j=0$ можно сопоставить кривую на проективной плоскости $\sum a_{ij}x^iy^jz^{n-i-j}=0$, где $n=\max(i+j)$.

Теперь уже можно проверить, правильно ли было наше предположение, что прямая $x=z/2$ и кривая $y^2z=x(x-z)(x-2z)$ пересекаются в бесконечно удаленной в направлении оси Oy точке. Подставив выражение $x=z/2$ в уравнение кривой, получим $y^2z=3z^3/8$. Это уравнение имеет решения трех видов: $z=0$, $y=kz$ и $y=-kz$, где $k=\sqrt{3/8}$. Следовательно, прямая $x=z/2$ на проективной плоскости пересекает рассматриваемую кривую действительно в трех точках: $(1/2,\sqrt{3/8},1)$, $(1/2,-\sqrt{3/8},1)$, $(0,1,0)$. Третья точка — бесконечно удаленная в направлении оси Oy. Можно даже нарисовать, как выглядят прямая $x=z/2$ и рассматриваемая кривая в окрестности бесконечно удаленной точки $(0,1,0)$. Для этого нужно вместо плоскости $z=1$ взять какую-нибудь плоскость, проходящую через точку $(0,1,0)$ и не проходящую через начало координат. Возьмем, например, плоскость $y=1$. На ней мы получим кривую $z=x(x-z)(x-2z)$. При малых x и z она выглядит почти как кривая $z=x^3$ (рис. 11).

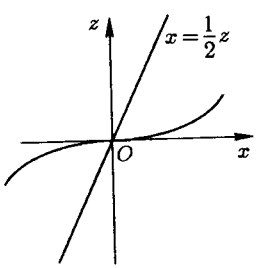


Рис. 11

Переход к проективной плоскости оказывается спасительным не только в разобранном выше случае. Любая прямая на проективной плоскости либо целиком принадлежит кубической кривой, либо пересекает ее с учетом кратности ровно в трех комплексных точках. Найти точки пересечения прямой $ax+by+cz=0$ и кубической кривой

$$\sum_{i+j+k=3} a_{ij}x^iy^jz^k=0$$

на проективной плоскости можно следующим образом. Одно из чисел a, b, c отлично от нуля. Пусть, например, $c\neq 0$. Тогда $z=\alpha x+\beta y$, где $\alpha=-a/c$ и $\beta=-b/c$ (случай $\alpha=\beta=0$ не исключается). Подставив это выражение в уравнение кубической кривой, получим уравнение вида

$$Q(x,y)=\sum b_p x^p y^{3-p}.$$

Возможны два случая.

1) Все коэффициенты b_p равны нулю. Тогда прямая $ax+by+cz=0$ целиком принадлежит кривой, т. е. Q делится на $ax+by+cz$.

2) Не все коэффициенты b_p равны нулю. Тогда

$$Q(x,y)=bx^ry^s(x-t_1y)\dots(x-t_my),$$

где $r+s+m=3$. Множителю x^r соответствует (r-кратная) точка пересечения $(0,1,\beta)$, множителю y^s — точка $(1,0,\alpha)$, множителю $x-t_iy$ — точка $(t_i,1,\alpha t_i+\beta)$.

У кубического многочлена Q с вещественными коэффициентами может быть либо три вещественных корня, либо один. Поэтому любая прямая на проективной плоскости пересекает с учетом кратности кубическую кривую либо в трех вещественных точках, либо в одной. Тем самым, со сложением несовпадающих точек кубической кривой мы почти разобрались. Неприятности остаются лишь для особых кривых с точками самопересечения или точками возврата. Дело в том, что любая прямая, проходящая через такую точку, имеет с кривой кратное пересечение (см. задачи 2 и 3). Поэтому при сложении такой точки с любой другой точкой мы должны были бы всегда получать один и тот же результат. Более подробно об особых кубических кривых мы поговорим в § 5.

Задачи.

1. Докажите, что кривая $y^2=x^3+px+q$ пересекает бесконечно удаленную прямую $z=0$ в одной точке, причем кратность пересечения равна трем.

2. Докажите, что в точке $(0,0)$ любая прямая пересекает кривую $y^2=x^2(x+1)$ с кратностью не меньше двух, а для прямых $y=\pm x$ кратность равна трем.

3. Докажите, что в точке $(0,0)$ любая прямая пересекает кривую $y^2 = x^3$ с кратностью не меньше двух, а для прямой $y = 0$ кратность равна трем.

4. Докажите, что на комплексной проективной плоскости любая окружность проходит через бесконечно удаленные точки $(1, i, 0)$ и $(1, -i, 0)$.

§ 3. Касательные и точки перегиба

Для сложения точек A и B, лежащих на кубической кривой, нужно провести прямую AB. А как следует поступить в том случае, когда точки A и B совпадают? Будем считать, что точка A неподвижна, а точка B движется к A по данной кривой. Тогда при определенных условиях (сформулированных ниже) прямая AB стремится к некоторой фиксированной прямой — касательной в точке A. Поэтому для нахождения суммы $A + A$ вместо прямой AB нужно провести касательную в точке A.

Если кривая, проходящая через точки A и B, задается уравнением $F = 0$, то ограничение F на прямую AB имеет корни в точках A и B. В предельном положении, когда точки A и B совпадают, ограничение F в точке A имеет кратный корень. Таким образом, ограничение F на касательную имеет кратный корень в точке касания. Это свойство можно использовать для того, чтобы получить уравнение касательной.

Пусть точка $P = (p_1, p_2, p_3)$ принадлежит кривой $F = 0$, т. е. $F(P) = 0$, а $X = (x_1, x_2, x_3)$ — произвольная точка. Точки проективной прямой PX имеют вид $\lambda P + \mu X$; отличные от X точки этой прямой имеют вид $P + tX$. Рассмотрим ограничение F на прямую PX как функцию t. В интересующем нас случае F есть многочлен степени три, поэтому

$$F(P + tX) = F(P) + at + bt^2 + ct^3 = Q(t),$$

где

$$F(P) = 0, \quad a = \sum F_i(P)\, x_i, \quad b = \frac{1}{2} \sum F_{ij}(P)\, x_i x_j$$

(F_i — производная F по i-й переменной). Точка P соответствует значению $t = 0$. Многочлен $Q(t)$ имеет в нуле кратный корень, если $a = 0$, т. е. $\sum F_i(P)\, x_i = 0$. Точка P, для которой хотя

бы одно из чисел $F_i(P)$ отлично от нуля называется *неособой* точкой кривой. Для неособой точки P уравнение $\sum F_i(P)\, x_i = 0$ однозначно определяет прямую l, касающуюся кривой в точке P.

Геометрически касательная к кривой определена однозначно, независимо от системы координат. При нашем подходе независимость определения касательной и особой точки от выбора системы координат пока не видна. Докажем инвариантность этих определений. Посмотрим, что происходит при замене координат (x_1, x_2, x_3) на (u_1, u_2, u_3), где $x_i = \sum a_{ij} u_j$. Пусть

$$G(u_1, u_2, u_3) = F(x_1(u), x_2(u), x_3(u)).$$

Тогда

$$G_j = \frac{\partial G}{\partial u_j} = \sum_i \frac{\partial F}{\partial x_i} \frac{\partial x_i}{\partial u_j} = \sum_i F_i\, a_{ij}.$$

Так как матрица $J = (a_{ij})$ невырожденная, то набор (G_1, G_2, G_3) ненулевой тогда и только тогда, когда набор (F_1, F_2, F_3) ненулевой. Если f и g — строки (F_1, F_2, F_3) и (G_1, G_2, G_3), x и u — столбцы (x_1, x_2, x_3) и (u_1, u_2, u_3), то $x = Ju$ и $g = fJ$. Поэтому $gu = (fJ)(J^{-1}x) = fx$, а значит, уравнения $fx = 0$ и $gu = 0$ определяют одну и ту же прямую.

Чтобы перейти от проективных координат (x_1, x_2, x_3) к обычным координатам (x_1, x_2), нужно положить $x_3 = 1$. Для соблюдения условия $x_3 = 1$ точки прямой PX нужно записать в виде $P + t(X - P)$. Разложение

$$F(P + t(X - P)) = \sum F_i(P)\,(x_i - p_i)\, t + \ldots$$

позволяет записать уравнение касательной в виде

$$F_1(P)\, x_1 + F_2(P)\, x_2 = F_1(P)\, p_1 + F_2(P)\, p_2.$$

Для проективных координат, т. е. в случае однородной функции F, выражение $\sum F_i(P)\, p_i$ равняется нулю. Дело в том, что для любого однородного многочлена F степени n справедлива *формула Эйлера*

$$\sum F_i(X)\, x_i = nF(X).$$

Эту формулу достаточно проверить для монома $x_1^r x_2^s x_3^t$, где $r + t + s = n$. Для положительных r, s и t производные этого монома

по x_1, x_2 и x_3 равны $rx_1^{r-1}x_2^s x_3^t$, $sx_1^r x_2^{s-1}x_3^t$ и $tx_1^r x_2^s x_3^{t-1}$. После домножения их соответственно на x_1, x_2 и x_3 получим $rx_1^r x_2^s x_3^t$, $sx_1^r x_2^s x_3^t$ и $tx_1^r x_2^s x_3^t$. Остается вспомнить, что $r+s+t=n$. Случай, когда среди чисел r, s и t есть нулевые, разбирается аналогично.

Пусть P — неособая точка кривой $F=0$. Тогда определена касательная l в точке P. Ограничение многочлена F на прямую l имеет в точке P кратный корень. Если кратность этого корня не меньше трех, то P называется *точкой перегиба*. Иными словами, из условия $a=\sum F_i(P)\,x_i=0$ должно следовать, что

$$b=\frac{1}{2}\sum F_{ij}(P)\,x_ix_j=0,$$

т. е. квадрика $\sum F_{ij}(P)\,x_ix_j=0$ должна содержать прямую $\sum F_i(P)\,x_i=0$.

Напомним, что многочлен второй степени $x^{\mathrm{T}}Ax$, записанный в матричной форме, делится на линейную функцию $x^{\mathrm{T}}l$ лишь в том случае, когда $x^{\mathrm{T}}Ax=x^{\mathrm{T}}lm^{\mathrm{T}}x$. Это означает, что матрица $A=lm^{\mathrm{T}}$ есть произведение столбца на строку, т. е. имеет ранг 1. В частности, $\det A=0$. Таким образом, если P — точка перегиба, то $\det(F_{ij}(P))=0$.

Покажем, что для неособой точки кривой верно и обратное, т. е. если P — неособая точка и $\det(F_{ij}(P))=0$, то P — точка перегиба. Рассмотрим квадрику

$$\sum F_{ij}(P)\,x_ix_j=0.$$

Точка P принадлежит ей, так как согласно формуле Эйлера

$$\sum F_{ij}(P)\,p_ip_j=2\sum F_j(P)\,p_j=6F(P)=0.$$

Кроме того, прямая $\sum F_i(P)\,x_i=0$ касается этой квадрики в точке P. В самом деле, уравнение касательной к квадрике

$$\sum F_{ij}(P)\,x_ix_j=0$$

в точке P имеет вид

$$\sum F_{ij}(P)\,x_ip_j=0,$$

а согласно формуле Эйлера

$$\sum F_{ij}(P)\,x_ip_j=2\sum F_i(P)\,x_i.$$

Поэтому касательная к кривой в точке P касается квадрики

$$\sum F_{ij}(P)\,x_ix_j=0,$$

которая вырождена в силу условия $\det(F_{ij}(P))=0$. Следовательно, квадрика состоит из пары прямых и целиком содержит касательную.

Подведем итоги. Точки пересечения кривых $F=0$ и $H=0$, где $H(X)=\det(F_{ij}(X))$, содержат все точки перегиба кривой $F=0$. Точками перегиба не будут лишь особые точки кривой $F=0$. Кривая $H=0$ называется *кривой Гессе*, или *гессианом*, кривой $F=0$. Если F — однородный многочлен степени n, то F_{ij} — однородный многочлен степени $n-2$. Поэтому H — однородный многочлен степени $3(n-2)$. Для кубического многочлена F многочлен H тоже кубический.

Инвариантность понятия точки перегиба и инвариантность кривой Гессе доказываются почти так же, как мы доказывали инвариантность касательной. Пусть

$$G(u_1,u_2,u_3)=F(x_1(u),x_2(u),x_3(u)),$$

где $x_i=\sum a_{ij}u_j$. Тогда

$$G_{pq}=\frac{\partial^2 G}{\partial u_p\partial u_q}=\sum_{i,j}\frac{\partial^2 F}{\partial x_i\partial x_j}\frac{\partial x_i}{\partial u_p}\frac{\partial x_j}{\partial u_q}=\sum_{i,j}a_{ip}F_{ij}a_{jq},$$

т. е. $(G_{pq})=J^{\mathrm{T}}(F_{ij})\,J$, где $J=(a_{ij})$. Следовательно, $\det(G_{pq})=(\det J)^2\det(F_{ij})$. Поэтому условия $\det(F_{ij})=0$ и $\det(G_{pq})=0$ эквивалентны.

Нахождение точек перегиба кривой сводится к нахождению точек ее пересечения с гессианом. Для этого требуется найти точки пересечения двух кривых. В этом случае, когда одна из кривых — прямая, мы это уже делали. Уравнение прямой позволяет выразить одну переменную через другую. Подставив это выражение в уравнение кривой, можно исключить одну переменную. Для кривых произвольной степени тоже можно исключить одну переменную, но сделать это несколько сложнее. Для большей наглядности мы сначала рассмотрим кривые в обычных координатах (x,y) и лишь затем перейдем к проективным координатам (x,y,z). Для простоты ограничимся случаем кривых третьей степени.

Многочлены $f(x,y)$ и $h(x,y)$ третьей степени можно записать в виде

$$f(x,y) = a_0y^3 + a_1(x)\,y^2 + a_2(x)\,y + a_3(x),$$

$$h(x,y) = b_0y^3 + b_1(x)\,y^2 + b_2(x)\,y + b_3(x),$$

где $a_k(x)$ и $b_k(x)$ — многочлены степени не более k, $0 \leqslant k \leqslant 3$. Если (x_0, y_0) — общая точка кривых $f(x,y) = 0$ и $h(x,y) = 0$, то многочлены

$$f_0(y) = a_0y^3 + a_1y^2 + a_2y + a_3$$

$$h_0(y) = b_0y^3 + b_1y^2 + b_2y + b_3,$$

где $a_k = a_k(x_0)$ и $b_k = b_k(x_0)$, рассматриваемые как многочлены от переменной y, имеют общий корень y_0. Верно и обратное: если f_0 и h_0 имеют общий корень y_0, то кривые $f(x,y) = 0$ и $h(x,y) = 0$ пересекаются в точке (x_0, y_0).

Над полем $\mathbb{C}$ у двух многочленов общий корень есть тогда и только тогда, когда у них есть общий делитель (над полем $\mathbb{R}$ общий делитель может не иметь корней). Если $a_0b_0 \neq 0$, то у многочленов $f_0(y)$ и $h_0(y)$ общий делитель есть тогда и только тогда, когда существуют такие многочлены h_1 и f_1, что $f_0h_1 = h_0f_1$, причем степени h_1 и f_1 меньше, чем степени h и f соответственно. В самом деле, если у f_0 и h_0 есть общий делитель d, то можно положить $f_1 = f_0d^{-1}$ и $h_1 = h_0d^{-1}$. А если $f_0h_1 = h_0f_1$, причем $\deg f_1 < \deg f_0$, то в hf_1 должны входить все простые множители, входящие в f_0, причем в тех же степенях. Войти же все в f_1 они не могут. Ограничение $a_0b_0 \neq 0$, безусловно, обременительное, но в проективном случае удовлетворить ему будет легко.

Пусть

$$h_1(y) = u_0y^2 + u_1y + u_2,$$

$$f_1(y) = v_0y^2 + v_1y + v_2.$$

Равенство $fh_1 = hf_1$ можно записать в виде

$$\begin{cases} a_0u_0 & -b_0v_0 & = 0,\\ a_1u_0+a_0u_1 & -b_1v_0-b_0v_1 & = 0,\\ a_2u_0+a_1u_1+a_0u_2 & -b_2v_0-b_1v_1-b_0v_2 & = 0,\\ a_3u_0+a_2u_1+a_1u_2 & -b_3v_0-b_2v_1-b_1v_2 & = 0,\\ a_3u_1+a_2u_2 & -b_3v_1-b_2v_2 & = 0,\\ a_3u_2 & -b_3v_2 & = 0. \end{cases}$$

Эта система линейных однородных уравнений относительно u и v имеет ненулевое решение тогда и только тогда, когда ее определитель равен нулю, т. е.

$$\begin{vmatrix} a_0 & a_1 & a_2 & a_3 & & \\ & a_0 & a_1 & a_2 & a_3 & \\ & & a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 & & \\ & b_0 & b_1 & b_2 & b_3 & \\ & & b_0 & b_1 & b_2 & b_3 \end{vmatrix} = 0. \tag{3.1}$$

Этот определитель называется *результантом* многочленов f и h. Коэффициенты a_k и b_k зависят от x, поэтому определитель (3.1) есть многочлен R от x, возможно нулевой. Для каждого корня x_0 многочлена $R(x)$ у кривых $f(x,y) = 0$ и $h(x,y) = 0$ есть общая точка (x_0, y_0). Отметим, что в вещественном случае при $x_0 \in \mathbb{R}$ число y_0 не обязательно вещественно. Если многочлен $R(x)$ нулевой, то у кривых есть общая компонента.

Повторим теперь предыдущие рассуждения для проективных координат, записав многочлены f и h в виде

$$f(x,y,z) = a_0z^3 + a_1(x,y)\,z^2 + a_2(x,y)\,z + a_3(x,y),$$

$$h(x,y,z) = b_0z^3 + b_1(x,y)\,z^2 + b_2(x,y)\,z + b_3(x,y),$$

где a_k и b_k однородные многочлены степени k, $0 \leqslant k \leqslant 3$. Координаты можно выбрать так, чтобы кривые $f = 0$ и $h = 0$ не проходили через точку $(0,0,1)$. Тогда будет выполняться нужное нам условие $a_0b_0 \neq 0$. Определитель (3.1) в проективном случае есть некоторый многочлен $R(x,y)$. Докажем, что он либо нулевой, либо однородный степени девять. В самом деле

$$R(\lambda x, \lambda y) = \begin{vmatrix} a_0 & \lambda a_1 & \lambda^2 a_2 & \lambda^3 a_3 & & \\ & a_0 & \lambda a_1 & \lambda^2 a_2 & \lambda^3 a_3 & \\ & & a_0 & \lambda a_1 & \lambda^2 a_2 & \lambda^3 a_3 \\ b_0 & \lambda b_1 & \lambda^2 b_2 & \lambda^3 b_3 & & \\ & b_0 & \lambda b_1 & \lambda^2 b_2 & \lambda^3 b_3 & \\ & & b_0 & \lambda b_1 & \lambda^2 b_2 & \lambda^3 b_3 \end{vmatrix}.$$

Домножим вторую и пятую строки на λ, а третью и шестую на λ^2. В результате получим определитель $R(x,y)$, в котором k-й столбец домножен на λ^k. Следовательно, $\lambda^6 R(\lambda x, \lambda y) = \lambda^{15} R(x,y)$, т. е. $R(\lambda x, \lambda y) = \lambda^9 R(x,y)$.

В общем случае, для многочленов степеней m и n степень многочлена $R(x,y)$ равна mn. Действительно,

$$\lambda^{p+q}R(\lambda x,\lambda y)=\lambda^r R(x,y),$$

где

$$p=1+2+\ldots+(n-1)=\frac{n(n-1)}{2},$$
$$q=\frac{m(m-1)}{2},$$
$$r=1+\ldots+(m+n-1)=\frac{(m+n)(m+n-1)}{2}.$$

Легко проверить, что $r-p-q=mn$.

Ненулевой многочлен $R(x,y)$ можно представить в виде $\prod_{i=1}^{9}(y_ix-x_iy)$, где числа x_i и y_i не обращаются в нуль одновременно. Для каждой из девяти пар (x_i,y_i) найдется такое число z_i, что (x_i,y_i,z_i) — точка пересечения кривых $f=0$ и $h=0$. Многочлен $R(x,y)$ может иметь кратные корни, т. е. некоторые пары (x_i,y_i) могут быть пропорциональными. Поэтому девять различных общих точек есть не у любой пары кубических кривых. Но на комплексной проективной плоскости одна общая точка есть всегда. Значит, у любой неособой кубической кривой есть точка перегиба. Именно это свойство нам понадобится в следующем параграфе.

Задачи

1. Докажите, что точка (x_0,y_0) кривой $y=f(x)$ является точкой перегиба тогда и только тогда, когда $f''(x_0)=0$.

2. Докажите, что все точки кривой

$$y^2=(x-x_1)(x-x_2)(x-x_3)$$

неособые тогда и только тогда, когда числа x_i попарно различны.

3. Докажите, что все точки кривой

$$y=(x-x_1)(x-x_2)(x-x_3),$$

кроме точки $(0,1,0)$, неособые.

4. Пусть A и B — точки перегиба кубической кривой, C — третья точка пересечения прямой AB с кубической кривой. Докажите, что C — точка перегиба.

Указание. Примените теорему 1.1 в случае, когда $p_1=p_2=p_3=AB$, q_1, q_2 и q_3 — касательные в точках A, B и C.

5. Касательные к кубической кривой в точках A и B пересекаются в точке перегиба P, прямая AB пересекает кривую в точке C. Докажите, что PC — касательная.

Указание. Примените теорему 1.1 в случае, когда p_1 — касательная в точке P, $p_2=p_3=AB$, q_1 и q_2 — касательные в точках A и B, $q_3=PC$.

§ 4. Нормальные формы неособой кубической кривой

Кубическая кривая называется *неособой,* если все ее точки неособые. В этом параграфе мы докажем, что над полем $\mathbb{C}$ уравнение неособой кубической кривой линейными заменами однородных координат можно привести к любому из следующих видов:

1) $y^2z=x^3+pxz^2+qz^3$ (форма Вейерштрасса),

2) $x^3+y^3+z^3=3\lambda xyz$.

При этом в первом случае многочлен x^3+px+q не имеет кратных корней (иначе кривая особая), а во втором случае $\lambda^3\neq 1$ (иначе кривая состоит из трех прямых).

Рассмотрим неособую кубическую кривую

$$\sum a_{ij}x^iy^jz^{3-i-j}=0$$

над полем $\mathbb{C}$. В предыдущем параграфе было доказано, что у нее есть точка перегиба. Можно считать, что точка перегиба имеет координаты $(0,1,0)$, причем касательная в этой точке задается уравнением $z=0$. Иными словами, ограничение функции

$$F(x,y,z)=\sum a_{ij}x^iy^jz^{3-i-j}$$

на прямую $z=0$, т. е. многочлен

$$a_{30}x^3+a_{21}x^2y+a_{12}xy^2+a_{03}y^3,$$

имеет корень $x = 0$ кратности три. Следовательно, $a_{21} = a_{12} = a_{03} = 0$, но $a_{30} \neq 0$, так как иначе рассматриваемая кривая целиком содержала бы прямую $z = 0$. Касательная в точке $(0,1,0)$ задается уравнением

$$F_x(0,1,0)\,x + F_y(0,1,0)\,y + F_z(0,1,0)\,z = 0.$$

Поэтому $F_x(0,1,0) = F_y(0,1,0) = 0$, но $F_z(0,1,0) \neq 0$, так как иначе точка $(0,1,0)$ была бы особой. Однородный многочлен $F_z(x,y,z)$ степени 2 при $x = z = 0$ и $y = 1$ принимает значение a_{02}. Можно считать, что $a_{02} = 1$. В обычных, не проективных координатах уравнение кривой запишется тогда в виде

$$y^2 - 2(ax+b)\,y + P_3(x) = 0,$$

где P_3 — многочлен третьей степени. Сделав замену $y_1 = y - ax - b$, получим $y_1^2 - (ax+b)^2 + P_3(x) = 0$, т. е. $y_1^2 = Q_3(x)$, где $Q_3(x) = (ax+b)^2 - P_3(x)$ — многочлен третьей степени. Заменой $x = \lambda x_1 + \mu$ его можно привести к виду $x_1^3 + px_1 + q$.

Многочлен Q_3 не имеет кратных корней, так как иначе уравнение кривой можно было бы привести к виду $y^2 = x^2(\alpha x + \beta)$, а для такой кривой начало координат есть особая точка.

В предыдущем параграфе было доказано, что с учетом кратности кубическая кривая имеет 9 точек перегиба. Но мы не смогли выяснить, все ли они различны. Если уравнение неособой кубической кривой записать в виде $y^2 = Q_3(x)$, то легко найти точки ее пересечения с гессианом и доказать, что все они различны.

Теорема 1. *Неособая кубическая кривая в* $\mathbb{C}P^2$ *имеет ровно* 9 *различных точек перегиба.*

Доказательство. Можно считать, что многочлен Q_3 имеет корень $x = 0$, т. е. рассматриваемая кривая задается уравнением $f(x,y) = 0$, где $f(x,y) = y^2 - x^3 - ax^2 - bx$. Так как у многочлена Q_3 нет кратных корней, то $b \neq 0$ и $a^2 - 4b \neq 0$. Чтобы получить уравнение кривой Гессе, перейдем к однородным координатам:

$$F(x,y,z) = y^2z - x^3 - ax^2z - bxz^2.$$

Тогда

$$H(x,y,z) = \begin{vmatrix} -6x - 2az & 0 & -2ax - 2bz \\ 0 & 2z & 2y \\ -2ax - 2bz & 2y & -2bx \end{vmatrix} = 8\left[(y^2 + bxz)(3x + az) - (ax + bz)^2\,z\right],$$

В обычных координатах уравнение кривой Гессе имеет вид

$$h(x,y) = y^2(3x + a) + bx(3x + a) - (ax + b)^2.$$

Найти точки пересечения кривых $f = 0$ и $h = 0$ можно легко. Запишем равенство $f = 0$ в виде $y^2 = x^3 + ax^2 + bx$ и подставим это выражение в уравнение $h = 0$. В результате получим

$$(x^3 + ax^2 + bx)(3x + a) + bx(3x + a) - (ax + b)^2 = 0,$$

т. е. $q(x) = 3x^4 + 4ax^3 + 6bx^2 - b^2 = 0$. Докажем, что многочлен $q(x)$ не имеет кратных корней. Его производная равна $12(x^3 + ax^2 + bx)$. Поэтому

$$q(x) - \frac{q'(x)}{12}\left(3x + a - \frac{b}{x}\right) = (4b - a^2)\,x^2.$$

Предположим, что $q(x_0) = q'(x_0) = 0$. Тогда $x_0 \neq 0$, так как $q(0) = -b^2 \neq 0$. С другой стороны $(4b - a^2)x_0^2 = 0$, причем $4b - a^2 \neq 0$. Поэтому $x_0 = 0$. Получено противоречие. Мы доказали, что многочлен $q(x)$ имеет четыре различных корня x_i.

Каждому корню x_i соответствуют два разных значения y, так как $y^2 = x_i^3 + ax_i^2 + bx_i = q'(x_i)/12 \neq 0$. Итак, кривые $F = 0$ и $H = 0$ имеют 8 точек пересечения в конечной области $z \neq 0$. А так как $F(x,y,0) = -x^3$ и $H(x,y,0) = 24xy^2$, то на прямой $z = 0$ кривые $F = 0$ и $H = 0$ имеют ровно одну общую точку $(0,1,0)$. □

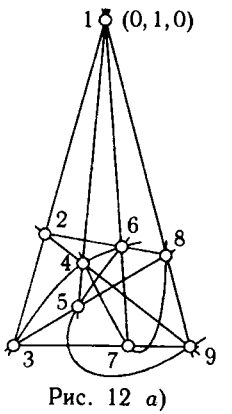


Рис. 12 а)

Любая прямая, проходящая через две точки перегиба, содержит еще одну точку перегиба. В самом деле, можно считать, что одна из точек перегиба имеет координаты $(0,1,0)$. Если (x_0,y_0) — общая точка кривой $y^2 = x^3 + ax^2 + bx$ и ее гессиана $y^2(3x + a) + bx(3x + a) = (ax + b)^2$, то $(x_0,-y_0)$ — тоже их общая точка. Точки $(0,1,0)$ и $(x_0,\pm y_0,1)$ лежат на прямой $x = x_0z$. Конфигурация из 9 точек перегиба и 12 содержащих их прямых схематично изображена на рис. 12, а). Более симметричная схема этой конфигурации изображена на рис. 12, б).

Теорема 2. *Заменой координат девять точек перегиба кубической кривой можно перевести в следующий набор точек:*

$$\begin{array}{lll}(0,1,-1), & (0,\varepsilon^2,-\varepsilon), & (0,\varepsilon,-\varepsilon^2),\\ (-1,0,1), & (-\varepsilon^2,0,1), & (-\varepsilon,0,1),\\ (1,-1,0), & (-\varepsilon,1,0), & (-\varepsilon^2,1,0),\end{array}$$

где $\varepsilon^3=1$ *и* $\varepsilon\neq 1$, *т. е.* $\varepsilon^2+\varepsilon+1=0$.

Доказательство. Прямые 189, 463, 527 (рис. 12, б) не могут пересекаться в одной точке. В самом деле, пусть R — общая точка этих прямых. Любые четыре точки общего положения в $\mathbb{CP}^2$ можно перевести проективным преобразованием в любые другие четыре точки общего положения. Поэтому можно считать, что точки R, 1, 5 и 6 вещественные. Но тогда вещественными будут и все остальные точки конфигурации. Легко убедиться, что это невозможно.

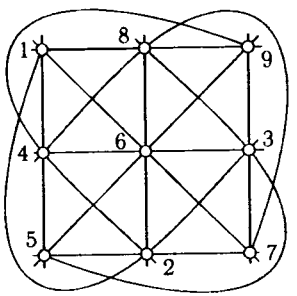


Рис. 12 *б*)

Добавив прямую 145 к указанной тройке прямых, получим четверку прямых общего положения. Поэтому можно считать, что прямые 145, 189, 463 и 527 задаются уравнениями $x+y+z=0$, $x=0$, $y=0$ и $z=0$ соответственно. Тогда координаты точек перегиба имеют следующий вид:

$$\begin{array}{lll}(0,1,-1), & (0,a,-b), & (0,c,-d),\\ (-1,0,1), & (-a',0,1), & (-c',0,1),\\ (1,-1,0), & (-b',1,0), & (-d',1,0),\end{array}\qquad(4.1)$$

причем все числа $a,b,\ldots,d'$ ненулевые.

Точки $(0,a,-b)$, $(-a',0,1)$, $(-b',1,0)$ лежат на одной прямой, а именно на прямой 862. Поэтому $ab'=ba'$. Можно считать, что $a=a'$ и $b=b'$. Аналогично $c=c'$ и $d=d'$.

Рассмотрев прямые 167 и 123, получим соответственно $a=d$ и $b=c$. Рассмотрев прямые 538 и 596, получим $a=bc$ и $c=ad$. Следовательно, $b^3=1$ и $a=b^2$. Ясно также, что $b\neq 1$. Подставив в (4.1) значения $a=\varepsilon^2$, $b=\varepsilon$, $c=\varepsilon$ и $d=\varepsilon^2$, получим требуемые точки.

Легко проверить, что остальные восемь прямых, содержащих тройки данных точек, имеют уравнения вида $x+\alpha y+\beta z=0$, где α и β принимают значения 1, ε и ε^2. □

С помощью теоремы 2 легко доказать, что любую неособую кубическую кривую можно привести к виду $x^3+y^3+z^3+3\lambda xyz=0$. В самом деле, пусть точки перегиба данной кривой имеют координаты, указанные в условии теоремы 2. Эти точки лежат как на тройке прямых $xyz=0$, так и на тройке прямых

$$(x+y+z)(x+\varepsilon y+\varepsilon^2 z)(x+\varepsilon^2 y+\varepsilon z)=0.$$

Поэтому любая кубическая кривая, проходящая через эти девять точек, задается уравнением

$$\mu xyz+\nu(x+y+z)(x+\varepsilon y+\varepsilon^2 z)(x+\varepsilon^2 y+\varepsilon z)=0.$$

Остается заметить, что

$$(x+y+z)(x+\varepsilon y+\varepsilon^2 z)(x+\varepsilon^2 y+\varepsilon z)=x^3+y^3+z^3-3xyz.$$

Задачи

1. Докажите, что линейной заменой $x'=ax+b$ кривую

$$y^2=(x-x_1)(x-x_2)(x-x_3)$$

можно привести к виду

$$y^2=x'(x'-1)(x'-\lambda),$$

где $\lambda=(x_3-x_1):(x_2-x_1)$. При этом для одной и той же кривой может получиться шесть разных значений λ, а именно, λ, λ^{-1}, $1-\lambda$, $(1-\lambda)^{-1}$, $(\lambda-1)\lambda^{-1}$, $\lambda(\lambda-1)^{-1}$.

2. Пусть $F=0$ — уравнение кубической кривой, $H=0$ — уравнение ее гессиана. Докажите, что $\lambda F+\mu H=0$ — уравнение кубической кривой с теми же точками перегиба, что и у исходной кривой (предполагается, что кривая невырожденная).

3. Докажите, что кривая $x^3+y^3+z^3=3\lambda xyz$ неособая тогда и только тогда, когда $\lambda^3\neq 1$.

4. Докажите, что кривая $x^3+y^2z+axz^2=0$ служит гессианом своего гессиана.

5. Докажите, что кривая $x^3 + y^3 + z^3 = 3\mu xyz$ служит гессианом кривой $x^3 + y^3 + z^3 = 3\lambda xyz$ тогда и только тогда, когда $\mu = -\dfrac{4+\lambda^3}{3\lambda^2}$.

6. Преобразовать кривую

$$y^2 = x(1-x)(1-\lambda^2 x),$$

полагая $\lambda^2 = \dfrac{4k}{(1+k)^2}$, $x = \dfrac{(1+k)^2 t}{(1+kt)^2}$, $y = \dfrac{(1+k)(1-kt)}{(1+kt)^3}$.

7. Преобразовать кривую

$$x^3 + y^3 - 3x - 3y + 4 = 0,$$

полагая $u = \dfrac{x-1}{y-1}$, $v = y - 1$.

8. Преобразовать кривую

$$x^3 + y^3 + 3(x^2 + y^2) + 2(x + y) + 1 = 0,$$

полагая $u = x + y$, $v = xy$.

9. Исследовать действие рационального преобразования

$$u = 1, \quad v = y/x$$

на кривую $x^2 - x^3 - y^2 = 0$. Сделать то же самое для кривой $x^2 - x^4 - y^2 = 0$.

§ 5. Особые кубические кривые

Уравнение неособой кубической кривой можно записать в виде

$$y^2 = (x - x_1)(x - x_2)(x - x_3),$$

где числа x_1, x_2 и x_3 попарно различны. В вещественном случае такая кривая изображена на рис. 13. Пусть $x_1 < x_2 < x_3$. При слиянии корней x_1 и x_2 получается кривая $y^2 = x^2(x-1)$ (рис. 14, а). При слиянии корней x_2 и x_3 получается кривая вида $y^2 = x^2(x+1)$ (рис. 14, б). Над полем $\mathbb{R}$ эти кривые различны, но над полем $\mathbb{C}$

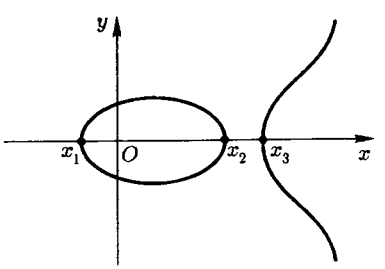


Рис. 13

их различие исчезает. При слиянии всех трех корней получается кривая $y^2 = x^3$ (рис. 14, в).

Для всех этих трех кривых начало координат — особая точка.

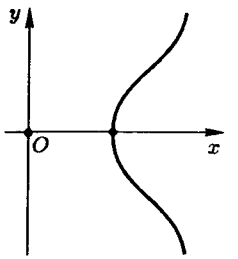


Рис. 14 *а*)

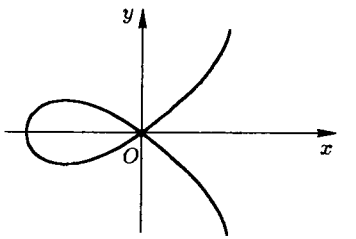


Рис. 14 *б*)

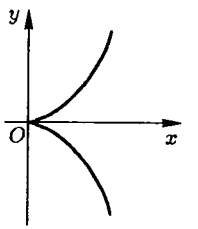


Рис. 14 *в*)

Любая прямая $y = kx$ пересекает кривые $y^2 = x^2(x \pm 1)$ и $y^2 = x^3$ в особой точке не менее чем двукратно. В самом деле, для уравнений $k^2x^2 = x^2(x \pm 1)$ и $k^2x^2 = x^3$ корень $x = 0$ не менее, чем двукратный. Поэтому для любой прямой, соединяющей особую точку с другой точкой кубической кривой, третьей точкой пересечения будет особая точка. Тем самым, сложение особой точки с любой другой точкой всегда должно приводить к одному и тому же результату. Поэтому для особой точки определить сложение не удастся. Но если особую точку исключить, то как для кривой $y^2 = x^2(x+1)$, так и для кривой $y^2 = x^3$ сложение точек определить можно.

Мы докажем, что если в качестве нулевого элемента в обоих случаях взять бесконечно удаленную точку, то кривая $y^2 = x^2(x+1)$ над полем $\mathbb{R}$ превращается в группу ненулевых действительных чисел по умножению, а кривая $y^2 = x^3$ — в группу действительных чисел по сложению.

Начнем с кривой $y^2 = x^3$. Она допускает рациональную параметризацию $x = t^{-2}$, $y = t^{-3}$. Точки пересечения этой кривой с прямой $ax + by + c = 0$ определяются соотношением

$$ct^3 + at + b = 0.$$

Если прямая не проходит через особую точку, то $c \neq 0$. В этом случае получаем кубическое уравнение с нулевым коэффициентом при t^2. Сумма корней такого уравнения равна нулю: $t_1 + t_2 + t_3 = 0$. Возьмем в качестве нулевого элемента E бесконечно удаленную точку, соответствующую параметру $t_E = 0$.

Пусть точкам A и B данной кривой соответствуют значения параметра t_A и t_B. Прямая AB пересекает кубическую кривую в точке X; при этом $t_A + t_B + t_X = 0$. Прямая EX пересекает кривую в точке $A + B$, т. е. $t_E + t_X + t_{A+B} = 0$. Следовательно, $t_{A+B} = -t_X = t_A + t_B$. При сложении точек кривой $y^2 = x^3$ складывается соответствующие им значения параметра t. Отметим, что особой точке соответствует значение параметра $t = \infty$.

Кривая $y^2 = x^2(x+1)$ также допускает рациональную параметризацию. В самом деле, пусть $y = tx$. Тогда $t^2x^2 = x^2(x+1)$, т. е. $x = t^2 - 1$ и $y = tx = t^3 - t$. Прямая $ax + by + c = 0$ пересекает кривую $y^2 = x^2(x+1)$ в точках, значения параметра для которых удовлетворяют соотношению

$$a(t^2 - 1) + b(t^3 - t) + c = 0.$$

Если $b \neq 0$, то после деления на b получим кубическое уравнение с коэффициентом 1 при t^3 и -1 при t. Корни такого уравнения удовлетворяют соотношению

$$t_1t_2 + t_2t_3 + t_3t_1 = -1.$$

Более простое соотношение можно получить после замены $t = (1+\tau)(1-\tau)^{-1}$. В самом деле, легко проверить, что $\tau_1\tau_2\tau_3 = 1$. Возьмем в качестве нулевого элемента E бесконечно удаленную точку, соответствующую значению параметра $\tau_E = 1$. Для нахождения суммы $A + B$ нужно рассмотреть точку X, в которой прямая AB пересекает кубическую кривую. Так как $\tau_A\tau_B\tau_X = 1$ и $\tau_X\tau_E\tau_{A+B} = 1$, то $\tau_{A+B} = \tau_A\tau_B$. При сложении точек кривой

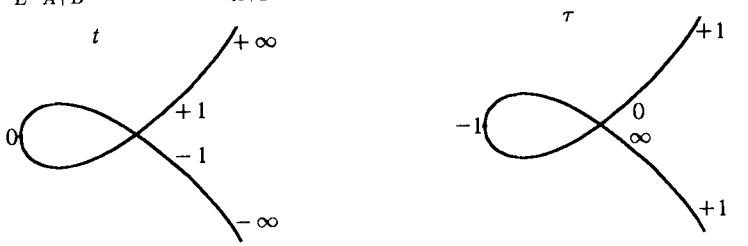


Рис. 15

$y^2 = x^2(x+1)$ соответствующие им значения параметра τ перемножаются. Особой точке соответствует не одно, а два значения каждого из параметров t и τ, а именно $t = \pm 1$ и $\tau = 0, \infty$ (рис. 15).

Задачи

1. Докажите, что точки $(t_1^{-2}, t_1^{-3}), \ldots, (t_6^{-2}, t_6^{-3})$, лежащие на кривой $x^3 = y^2$, принадлежат одной кривой второго порядка тогда и только тогда, когда $t_1 + \ldots + t_6 = 0$.

2. Кривая $y^2 = x^2(x-1)$ имеет рациональную параметризацию $x = t^2 + 1$, $y = t^3 + t$. Докажите, что точки, соответствующие значениям параметра t_1, t_2 и t_3, лежат на одной прямой тогда и только тогда, когда $t_1t_2 + t_2t_3 + t_3t_1 = 1$.

3. Неприводимая кривая третьего порядка, имеющая особую точку с несовпадающими касательными, имеет три точки перегиба, лежащие на одной прямой. Ее уравнение можно привести к виду $y^2 = x^2(x-1)$.

4. Неприводимая кривая третьего порядка, имеющая особую точку с совпадающими касательными, имеет одну точку перегиба. Ее уравнение можно привести к виду $y^2 = x^3$.

§ 6. Неособая кубическая кривая не допускает рациональной параметризации

Особые кубические кривые, которыми мы занимались в предыдущем параграфе, допускают рациональную параметризацию. Но никакая неособая кубическая кривая такой параметризации не допускает. Напомним, что неособую кубическую кривую можно привести к виду $y^2 = x(x-1)(x-\lambda)$, где $\lambda \neq 0, 1$.

Теорема 1. *Если $\lambda \neq 0, 1$, то не существует таких многочленов P_1, P_2, Q_1, Q_2, что функции $y(t) = P_1(t)/P_2(t)$ и $x(t) = Q_1(t)/Q_2(t)$ не постоянные и удовлетворяют соотношению $y^2 = x(x-1)(x-\lambda)$.*

Доказательство. Предположим, что $P_1(t)/P_2(t)$ и $Q_1(t)/Q_2(t)$ — не константы, причем

$$\frac{P_1^2}{P_2^2} = \frac{Q_1}{Q_2}\,\frac{Q_1 - Q_2}{Q_2}\,\frac{Q_1 - \lambda Q_2}{Q_2}.$$

Тогда можно считать, что многочлены P_1 и P_2 взаимно просты, Q_1 и Q_2 тоже взаимно просты. Так как

$$P_1^2Q_2^3 = P_2^2Q_1(Q_1 - Q_2)(Q_1 - \lambda Q_2),$$

то многочлен P_2^2, который взаимно прост с P_1^2, делится на Q_2^3, а многочлен Q_2^3, который взаимно прост с Q_1, $Q_1 - Q_2$ и $Q_1 - \lambda Q_2$, делится на P_2^2. Следовательно, многочлены Q_2^3 и P_2^2 пропорциональны. Поэтому после замены многочлена P_1 на пропорциональный ему многочлен можно получить равенство

$$P_1^2 = Q_1 (Q_1 - Q_2)(Q_1 - \lambda Q_2). \qquad (6.1)$$

Кроме того, многочлен Q_2^3 — квадрат некоторого многочлена, а значит, Q_2 тоже квадрат некоторого многочлена.

Многочлены Q_1, $Q_1 - Q_2$ и $Q_1 - \lambda Q_2$ попарно взаимно просты, поэтому из равенства (6.1) следует, что каждый из них — полный квадрат. Итак, в семействе многочленов вида $\alpha Q_1 + \beta Q_2$, где $\alpha, \beta \in \mathbb{C}$, нашлись четыре полных квадрата, а именно, Q_1, Q_2, $Q_1 - Q_2$, $Q_1 - \lambda Q_2$. Эти многочлены не пропорциональны и попарно различны, поскольку $\lambda \neq 0, 1$.

Чтобы прийти к противоречию, покажем, что на проективной прямой $\alpha Q_1 + \beta Q_2$, где многочлены Q_1 и Q_2 взаимно просты, $\alpha, \beta \in \mathbb{C}$ и не обращаются в нуль одновременно, не более трех точек могут быть полными квадратами. Предположим, что на этой проективной прямой нашлись четыре полных квадрата. Два из них обозначим через R_1^2 и R_2^2. Так как R_1^2 не пропорционально R_2^2, то два оставшихся полных квадрата можно представить в виде $\alpha_1 R_1^2 - \beta_1 R_2^2$ и $\alpha_2 R_1^2 - \beta_2 R_2^2$. Многочлены R_1 и R_2 взаимно просты и поэтому многочлены $\sqrt{\alpha_i} R_1 \pm \sqrt{\beta_i} R_2$ должны быть полными квадратами, причем не пропорциональными. Таким образом среди многочленов $\alpha R_1 + \beta R_2$, где R_1 и R_2 — взаимно просты, $\alpha, \beta \in \mathbb{C}$ и не обращаются в нуль одновременно, существуют четыре попарно не пропорциональных квадрата. В результате от проективной прямой $\alpha Q_1 + \beta Q_2$, на которой есть четыре полных квадрата, мы переходим к проективной прямой $\alpha R_1 + \beta R_2$, на которой тоже есть четыре полных квадрата. От полученной проективной прямой можно перейти к следующей и т. д. Но при каждом таком переходе максимальная степень многочлена вида $\alpha Q_1 + \beta Q_2$ уменьшается по крайней мере в два раза. Получено противоречие. □

Задача

Приведите пример взаимно простых многочленов Q_1 и Q_2, для которых многочлены Q_1, $Q_1 + Q_2$ и $Q_1 + 2Q_2$ являются полными квадратами.

ГЛАВА 2

ЭЛЛИПТИЧЕСКИЕ ФУНКЦИИ

Сложение точек окружности связано с ее параметризацией функциями синус и косинус. В самом деле, рассмотрим отображение $f\colon \mathbb{R} \to S^1$, заданное формулой $f(t) = (\cos t, \sin t)$. Это отображение параметризует окружность действительными числами таким образом, что сложение точек окружности соответствует сложению действительных чисел.

Аналогичная параметризация существует и для кубических кривых. Она получается с помощью эллиптических функций. При этой параметризации сложение точек кубической кривой, определенное в гл. 1, соответствует сложению значений параметра.

В этой главе мы изучим основные свойства эллиптических функций и покажем, как с их помощью можно параметризовать неособую кубическую кривую.

Своим названием эллиптические функции обязаны эллипсу, хотя связь эта косвенная. Связующим звеном между ними послужили эллиптические интегралы. Дело в том, что длина дуги эллипса выражается эллиптическим интегралом некоторого частного вида. Именно отсюда и произошло название — «эллиптический интеграл». А эллиптические функции возникли при обращении эллиптических интегралов другого частного вида, никак не связанных с длиной дуги эллипса.

Эллиптические интегралы появились еще в XVII в. при вычислении длин дуг некоторых кривых. Помимо эллипса, важным примером для развития этой теории послужила лемниската Бернулли, длина дуги которой выражается интегралом вида $\int_0^\alpha \frac{dx}{\sqrt{1 - x^4}}$. Для этого интеграла в первой половине XVIII в.

итальянский математик граф Фаньяно получил теорему сложения

$$\int_0^{\alpha} \frac{dx}{\sqrt{1-x^4}} + \int_0^{\beta} \frac{dx}{\sqrt{1-x^4}} = \int_0^{\gamma} \frac{dx}{\sqrt{1-x^4}},$$

где

$$\gamma = \frac{\alpha\sqrt{1-\beta^4} + \beta\sqrt{1-\alpha^4}}{1+\alpha^2\beta^2}.$$

Исследования Фаньяно заинтересовали Эйлера. Ему удалось получить теорему сложения для интегралов более общего вида

$$\int_0^{\alpha} \frac{dx}{\sqrt{P(x)}},$$

где $P(x) = 1 + mx^2 + nx^4$. При этом

$$\int_0^{\alpha} \frac{dx}{\sqrt{P(x)}} + \int_0^{\beta} \frac{dx}{\sqrt{P(x)}} = \int_0^{\gamma} \frac{dx}{\sqrt{P(x)}},$$

где

$$\gamma = \frac{\alpha\sqrt{P(\beta)} + \beta\sqrt{P(\alpha)}}{1 - n\alpha^2\beta^2}.$$

Эйлер получил также теоремы сложения более общего вида.

Над развитием теории эллиптических интегралов много лет неутомимо трудился Лежандр. Свои результаты он изложил в книге «Exercises de calcul intégral» («Упражнения по интегральному исчислению»), опубликованной в 1811–1819 гг. Переработанное издание этой книги вышло в 1827–1832 гг. под названием «Traité des fonctions elliptiques et des intégrales eulériennes» («Трактат об эллиптических функциях и эйлеровых интегралах») [A7]. Эти три тома большого формата содержат огромное количество теорем о свойствах эллиптических интегралов и их приложениях. Эллиптическими функциями Лежандр называл то, что сейчас называют эллиптическими интегралами. После появления работ Абеля и Якоби значение книги Лежандра померкло. Но сами Абель и Якоби о книге Лежандра отзывались с большим уважением, как она того и заслуживает.

Собственно теория эллиптических функций начинается с работы Абеля «Recherches sur les fonctions elliptiques» («Исследования по эллиптическим функциям») [A1], которая была опубликована в 1827–1828 гг. в журнале Крелля. Абель показал, что при обращении эллиптического интеграла первого рода

$$\alpha = \int \frac{dx}{\sqrt{(1-cx^2)(1+ex^2)}}$$

появляется функция $\varphi(\alpha)$, имеющая в комплексной области два периода. Он детально исследовал уравнения, связывающие $\varphi(\alpha)$ и $\varphi(n\alpha)$. Почти одновременно с Абелем теорией эллиптических функций начал заниматься Якоби [A15]. Это привело к напряженному, хотя и недолгому состязанию между ними. Не имея постоянной работы, почти в нищете Абель завершил вторую часть «Recherches ...» и продолжил интенсивные исследования. Но вскоре он тяжело заболел и умер в 1829 г. в возрасте 27 лет.

Многое из того, что было обнаружено Абелем и Якоби, задолго до них знал Гаусс, но он не опубликовал своих результатов.

§ 1. Топологическое строение неособой кубической кривой в $\mathbb{C}P^2$

Эллиптические функции удобно рассматривать как функции комплексного переменного. Многие их свойства выявляются лишь на комплексной плоскости, а не на вещественной прямой. Параметризация кубической кривой тоже более наглядна над полем $\mathbb{C}$. Поэтому мы начнем с того, что выясним, как устроена топологически неособая кубическая кривая в $\mathbb{C}P^2$. Оказывается, что с топологической точки зрения все такие кривые устроены одинаково; все они — двумерные торы.

Уравнение любой неособой кубической кривой в $\mathbb{C}P^2$ можно привести к виду

$$y^2z = (x - a_1z)(x - a_2z)(x - a_3z), \qquad (1.1)$$

где числа a_i попарно различны. Это уравнение определяет в $\mathbb{C}P^2$ комплексную кривую, комплексная размерность которой равна 1, а вещественная равна 2.

Чтобы выяснить топологическое строение кривой (1.1) в $\mathbb{C}P^2$, рассмотрим проекцию p: $\mathbb{C}P^2 \setminus (0,1,0) \to \mathbb{C}P^1$, при которой

точка (x, y, z) переходит в точку (x, z). Комплексная проективная прямая $\mathbb{C}P^1$ (одноточечная компактификация $\mathbb{C}$) гомеоморфна двумерной сфере S^2. При $b \neq 0$ уравнение $y^2 = b$ имеет ровно два различных решения. Поэтому если $z \neq 0$ и $x - a_i z \neq 0$, то у точки $(x, z) \in \mathbb{C}P^1$ есть ровно два прообраза, лежащих на кривой (1.1). Если же $z \neq 0$, но x/z равно одному из чисел a_i, то прообраз один. В случае, когда $z = 0$, уравнение (1.1) превращается в уравнение $x^3 = 0$. Следовательно, у точки $\infty = (1, 0)$ прообраз тоже один, а именно, точка $(0, 1, 0)$. Точнее говоря, при стремлении z к нулю прообраз точки $(1, z)$ стремится к точке $(0, 1, 0)$.

Проекция p кривой (1.1) на $\mathbb{C}P^1$ устроена следующим образом. Если из $\mathbb{C}P^1$ исключить точки a_1, a_2, a_3 и ∞, то у всех остальных точек будет ровно по два прообраза. Строение отображения в окрестностях точек a_i и ∞ нужно изучить более детально. Будем для простоты считать, что $a_1 = 0$. Рассмотрим аффинные координаты, т. е. положим $z = 1$. Тогда (1.1) перепишется в виде

$$y^2 = x(x - a_2)(x - a_3),$$

где $a_2 a_3 \neq 0$. Величина $(x - a_2)(x - a_3)$ почти постоянна для точек x, близких к нулю, т. е. мы фактически имеем уравнение $y^2 = cx$. Это уравнение имеет решения вида $x = c\lambda^2 e^{2i\varphi}$, $y = c\lambda e^{i\varphi}$. Обозначим через (x_0, y_0) решение, отвечающее $\varphi = 0$. При изменении φ от 0 до π получим полный обход вокруг точки $(0, 1)$ на $\mathbb{C}P^1$, который начинается и заканчивается в точке $(x_0, 1)$. У величины y_0 при таком обходе изменяется знак. Проекция кривой (1.1) на $\mathbb{C}P^1$ в рассматриваемой системе координат устроена как $(x, y) \longmapsto x$. Поднимая на кривую (1.1) обход вокруг точки (0,1) на $\mathbb{C}P^1$, мы попадаем не в исходную точку (x_0, y_0), а точку $(x_0, -y_0)$ (рис. 14). Но повторив обход еще раз, мы возвратимся в исходную точку, так как, изменив дважды знак числа y_0, снова получим исходное число. Строение проекции кривой (1.1) на $\mathbb{C}P^1$ в окрестности точки ∞ такое же, как и в окрестностях

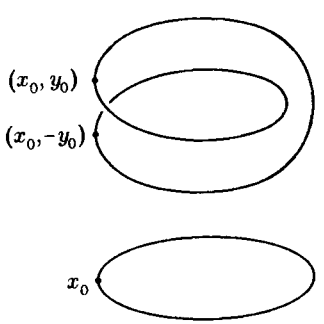


Рис. 14

точек a_i. В самом деле, пусть $x = 1$. Тогда в окрестности точки $z = 0$ уравнение (1.1) выглядит приблизительно как $y^2 = 1/z$. При полном обходе вокруг точки $z = 0$ величина y изменяет знак.

Сделаем на $\mathbb{C}P^1$ разрезы от a_1 до a_2 и от a_3 до ∞. Поднятия этих разрезов на кривую (1.1) разбивают ее на две части. В самом деле, при обходе вдоль любого замкнутого пути в $\mathbb{C}P^1$, не пересекающего разрезы, точки a_1, a_2, a_3 и ∞ мы будем обходить лишь парами, а при обходе вокруг двух точек величина y не изменяется. Поэтому из одного прообраза некоторой точки $\mathbb{C}P^1$ нельзя попасть во второй ее прообраз, не пересекая разрезов.

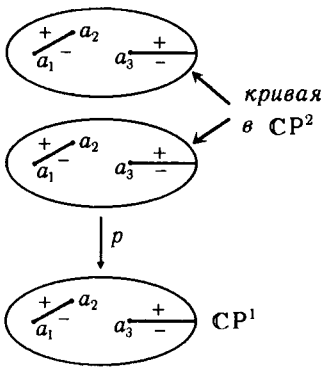


Рис. 15

Если на $\mathbb{C}P^1$ сделать разрезы от a_1 до a_2 и от a_3 до ∞, то оставшуюся часть $\mathbb{C}P^1$ можно представить в виде плоскости с разрезами (рис. 15).

Лежащая над ней часть кривой (1.1) состоит из двух кусков. Нужно лишь понять, как эти куски склеены. Пересекая в $\mathbb{C}P^1$ разрез, мы попадаем с края со знаком плюс одного куска кривой (1.1) на край со знаком минус другого куска. Произведя такую склейку краев, в результате получим тор (рис. 16).

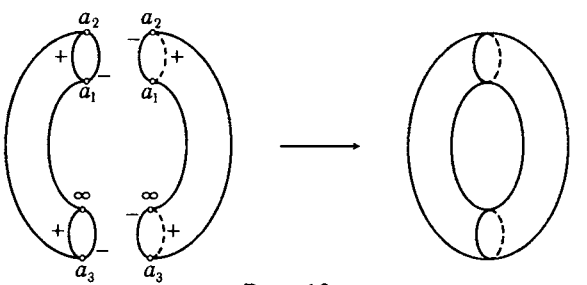


Рис. 16

Алгебраическую кривую называют *эллиптической*, если она как одномерное комплексное многообразие представляет собой тор.

Можно показать, что любая эллиптическая кривая биголоморфна некоторой неособой кубической кривой в $\mathbb{C}P^2$.

Параметризацию кубической кривой в $\mathbb{C}P^2$ можно задать с помощью отображения $f\colon \mathbb{C}^1 \longrightarrow \mathbb{C}P^2$, где $f(z) = (F_1(z), F_2(z), 1)$. Образ этого отображения должен быть тором. Простейшее отображение $\mathbb{C}^1$ на тор получается при отождествлении всех точек вида $z + n\omega_1 + m\omega_2$. Это соответствует тому, что ω_1 и ω_2 — периоды функций F_1 и F_2.

§ 2. Эллиптические функции

Функция f называется *двоякопериодической*, если существуют такие комплексные числа ω_1 и ω_2 (причем $\omega_1/\omega_2 \notin \mathbb{R}$), что для любых целых чисел m и n выполнено равенство $f(z + m\omega_1 + n\omega_2) = f(z)$. Для определенности будем считать, что $\operatorname{Im}(\omega_1/\omega_2) > 0$. Это означает, что поворот от ω_1 к ω_2 происходит по часовой стрелке (рис. 17).

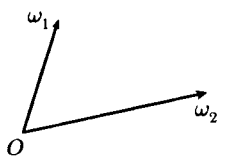


Рис. 17

В дальнейшем нас будут интересовать лишь мероморфные двоякопериодические функции. Напомним, что аналитическая функция называется *мероморфной*, если у нее в конечной области $\mathbb{C}$ нет особых точек, отличных от полюсов. В окрестности любой конечной точки a мероморфная функция f допускает разложение

$$f(z) = c_0(z-a)^r + c_1(z-a)^{r+1} + \ldots,$$

где $c_0 \neq 0$, $r \in \mathbb{Z}$. Мероморфная двоякопериодическая функция называется *эллиптической*.

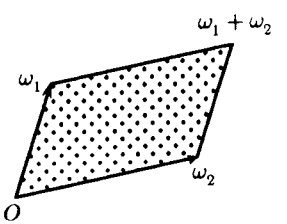


Рис. 18

Любое комплексное число z можно представить в виде $z = a_1\omega_1 + a_2\omega_2$, где $a_i \in \mathbb{R}$. Число a_i можно представить в виде суммы его целой и дробной частей, поэтому эллиптическая функция полностью определяется своими значениями в области

$$\{\alpha_1\omega_1 + \alpha_2\omega_2 \mid 0 \leqslant \alpha_1, \alpha_2 \leqslant 1\},$$

которая называется *фундаментальным параллелограммом* (рис. 18).

Теорема 1. *Эллиптическая функция, не имеющая полюсов, постоянна.*

Доказательство. Предположим, что эллиптическая функция $f(z)$ не имеет полюсов. Тогда функция $|f(z)|$ непрерывна на $\mathbb{C}$. А так как фундаментальный параллелограмм компактен, то $|f(z)| \leqslant M$ для некоторого числа M. Но тогда $|f(z)| \leqslant M$ для всех $z \in \mathbb{C}$. Итак, f — ограниченная целая функция на $\mathbb{C}$. Согласно теореме Лиувилля функция f постоянна. □

Все особые точки мероморфной функции, расположенные в конечной области, изолированные. Поэтому фундаментальный параллелограмм содержит лишь конечное число особых точек. Следовательно, параллелограмм можно параллельно сдвинуть так, что на его сторонах не будет особых точек. Образ фундаментального параллелограмма при параллельном переносе мы также будем называть фундаментальным параллелограммом, и в дальнейшем будем считать, что на его сторонах нет особых точек.

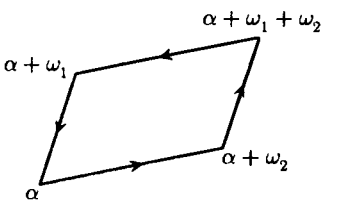


Рис. 19

Пусть P — фундаментальный параллелограмм с вершинами α, $\alpha + \omega_1$, $\alpha + \omega_1 + \omega_2$ и $\alpha + \omega_2$, ∂P — его граница (рис. 19).

Тогда $\displaystyle\int_{\partial P} f(z)\,dz = 0$ для любой эллиптической функции $f(z)$.

В самом деле, в этот интеграл входят интегралы

$$\int_{\alpha}^{\alpha+\omega_1} f(z)\,dz \qquad \text{и} \qquad \int_{\alpha+\omega_2}^{\alpha+\omega_1+\omega_2} f(z)\,dz$$

со знаками плюс и минус соответственно. Эти интегралы равны, так как $f(z + \omega_2) = f(z)$. Аналогичным образом сокращаются интегралы по другой паре сторон.

Из этого утверждения можно получить существенную информацию о нулях и полюсах эллиптической функции.

Теорема 2. *Справедливы утверждения:*

а) сумма вычетов особых точек эллиптической функции, расположенных внутри фундаментального параллелограмма, равна нулю;

б) пусть a_i — нули и полюсы эллиптической функции, расположенные внутри фундаментального параллелограмма, r_i — их порядки (порядки нулей положительные, а порядки полюсов отрицательные). Тогда $\sum r_i = 0$ и $\sum r_i a_i = m\omega_1 + n\omega_2$, где m и n — целые числа.

Доказательство. Если на границе фундаментального параллелограмма P нет особых точек функции g, то

$$\sum_P \operatorname{res} g = \frac{1}{2\pi i} \int_{\partial P} g(z)\,dz.$$

Для доказательства теоремы а) достаточно воспользоваться этим равенством, положив $g = f$.

Для доказательства равенства $\sum r_i = 0$ нужно положить $g(z) = f'(z)/f(z)$. Если $f(z)$ — эллиптическая функция, то функция $g(z) = f'(z)/f(z)$ тоже эллиптическая. Кроме того, если

$$f(z) = c_0(z-a)^r + c_1(z-a)^{r+1} + \dots,$$

то

$$g(z) = r(z-a)^{-1} + \alpha_1 + \alpha_2(z-a) + \dots,$$

а значит, вычет функции $g(z)$ в точке a равен r. Таким образом, $\sum r_i = 0$.

Для доказательства равенства $\sum r_i a_i = m\omega_1 + n\omega_2$ положим $g(z) = zf'(z)/f(z)$. Нам придется проделать некоторые вычисления, потому что функция $g(z) = zf'(z)/f(z)$ не обязательно эллиптическая. Предварительно заметим, что если

$$f(z) = c_0(z-a)^r + c_1(z-a)^{r+1} + \dots,$$

то

$$g(z) = \frac{a + (z-a)}{z-a} \frac{rc_0(z-a)^{r-1} + \dots}{c_0(z-a)^{r-1}} = ar(z-a)^{-1} + \dots,$$

а значит, вычет функции $g(z)$ в точке a равен ar. Теперь займемся вычислением интеграла $\int_{\partial P} g(z)\,dz$. В него входит разность интегралов

$$\int_\alpha^{\alpha+\omega_1} \frac{zf'(z)}{f(z)}\,dz \quad \text{и} \quad \int_{\alpha+\omega_2}^{\alpha+\omega_1+\omega_2} \frac{zf'(z)}{f(z)}\,dz = \int_\alpha^{\alpha+\omega_1} \frac{(z+\omega_2)f'(z)}{f(z)}\,dz.$$

Нетрудно проверить, что эта разность равна

$$-\omega_2 \int_\alpha^{\alpha+\omega_1} \frac{f'(z)}{f(z)}\,dz = -\omega_2 \ln f(z)\Big|_\alpha^{\alpha+\omega_1}.$$

А так как $f(\alpha+\omega_1) = f(\alpha)$, то при изменении z от α до $\alpha+\omega_1$ логарифм $f(z)$ может измениться лишь на $2k\pi i$. В итоге получаем, что в интеграл $\frac{1}{2\pi i}\int_{\partial P} g(z)\,dz$ одна пара сторон параллелограмма дает вклад $n\omega_2$, где n — целое число. Другая пара сторон дает вклад $m\omega_1$. □

Как мы уже говорили, непостоянная эллиптическая функция должна иметь хотя бы один полюс внутри фундаментального параллелограмма. Но так как сумма вычетов особых точек, лежащих внутри фундаментального параллелограмма, равна нулю, то функция не может иметь там ровно один плюс порядка -1. Число полюсов эллиптической функции, лежащих внутри фундаментального параллелограмма, с учетом их кратностей называют *порядком эллиптической функции*. Минимальный возможный порядок равен двум, при этом есть два варианта:

1) один полюс порядка -2 (так устроена функция Вейерштрасса, о которой пойдет речь в следующем параграфе);

2) два простых полюса (так устроены эллиптические функции Якоби, о которых рассказывается в § 8).

Согласно теореме 2, б) для эллиптической функции сумма порядков нулей, лежащих внутри фундаментального параллелограмма, равна сумме порядков полюсов, т. е. равна ее порядку. Ясно также, что у функции $f(z) - c$ полюсы те же самые, что и у функции $f(z)$. Поэтому эллиптическая функция порядка r любое конечное значение принимает внутри фундаментального параллелограмма с учетом кратности ровно r раз.

Задачи

1. Эллиптические функции f и g имеют одинаковые периоды, причем в каждом полюсе они имеют одинаковые главные части

$$c_r(z-a)^r + c_{r+1}(z-a)^{r+1} + \dots + c_{-1}(z-a)^{-1}.$$

Докажите, что разность этих функций постоянна.

2. Эллиптические функции f и g имеют одинаковые периоды, причем их нули и полюсы совпадают с учетом их кратности. Докажите, что отношение этих функций постоянно.

§ 3. Функция Вейерштрасса

Мы уже доказали некоторые свойства эллиптических функций, но еще не убедились, что множество эллиптических функций состоит не только из констант. Пора привести пример нетривиальной эллиптической функции. Докажем, что для любой решетки $\Lambda = \{m\omega_1 + n\omega_2 \mid m, n \in \mathbb{Z}\}$, $\omega_1, \omega_2 \in \mathbb{C}$, $\operatorname{Im}\omega_1/\omega_2 > 0$, эллиптической будет функция

$$\wp(z) = \frac{1}{z^2} + \sum{}' \left[\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right], \qquad (3.1)$$

где штрих у знака суммы означает, что суммирование ведется по всем ненулевым элементам $\omega \in \Lambda$. Группировка членов в квадратных скобках существенна, потому что по отдельности ряды $\sum'(z-\omega)^{-2}$ и $\sum'\omega^{-2}$ расходятся.

Докажем сначала, что ряд (3.1) действительно определяет мероморфную функцию. На любом компакте K, не содержащем точек решетки, этот ряд сходится равномерно и абсолютно. В самом деле,

$$\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} = \frac{2z\omega - z^2}{\omega^2(z-\omega)^2} = \frac{\omega}{\omega^4}\,\frac{2z - z^2\omega^{-1}}{(z\omega^{-1}-1)^2}.$$

Если число $|\omega|$ достаточно велико, то $\dfrac{2z - z^2\omega^{-1}}{(z\omega^{-1}-1)^2} \approx 2z$. Поэтому для всех $\omega \in \Lambda'$ с достаточно большим значением $|\omega|$ и для всех $z \in K$ найдется такая константа C, что

$$\left|\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right| < \frac{C}{|\omega|^3}.$$

Кроме того, $|z-\omega| > \varepsilon$ для всех $z \in K$ и $\omega \in \Lambda'$, поэтому такая константа найдется и для всех $\omega \in \Lambda'$. Легко проверить, что ряд $\sum'|\omega|^{-3}$ сходится. В самом деле,

$$\sum{}'|\omega|^{-3} = \sum_{n=1}^{\infty} \sum_{\max(|p|,|q|)=n} |p\omega_1 + q\omega_2|^{-3} \leqslant \sum_{n=1}^{\infty} 8n(nh)^{-3},$$

где h — наименьшая из высот фундаментального параллелограмма. Таким образом, $\wp(z)$ — мероморфная функция с полюсами в узлах решетки. Она называется *функцией Вейерштрасса*.

Перейдем к доказательству периодичности функции $\wp(z)$. Рассмотрим для этого ее производную

$$\wp'(z) = -2\sum(z-\omega)^{-3}$$

(здесь суммирование ведется уже по всем узлам решетки). Очевидно, что ω_1 и ω_2 — периоды функции $\wp'(z)$. Поэтому функции $\wp(z+\omega_i)$ и $\wp(z)$ могут отличаться лишь на константу c. Подставив значение $z = -\omega_i/2$ в равенство $\wp(z+\omega_i) = \wp(z) + c$, получим $\wp(\omega_i/2) = \wp(-\omega_i/2) + c$. Но из формулы (3.1) видно, что функция $\wp(z)$ четная. Поэтому $c = 0$, т. е. ω_1 и ω_2 — периоды функции $\wp(z)$.

Функция $\wp$ имеет в узлах решетки двукратные полюсы, других особых точек у нее нет. Внутри фундаментального параллелограмма расположен ровно один узел решетки. Поэтому сумма полюсов функции $\wp$, расположенных внутри фундаментального параллелограмма, сравнима с нулем по модулю Λ. Согласно теореме 2 § 2 внутри фундаментального параллелограмма расположены два нуля u и v функции $\wp$, причем $u+v \equiv 0 \pmod{\Lambda}$. Для любой константы c полюсы функции $\wp(z) - c$ совпадают с полюсами функции $\wp(z)$, поэтому внутри фундаментального параллелограмма есть ровно две точки u и v, для которых $\wp(u) = \wp(v) = c$, причем $u + v \equiv 0 \pmod{\Lambda}$. В том случае, когда $u \equiv -u \pmod{\Lambda}$, эти две точки совпадают, т. е. соответствующее значение функция $\wp$ принимает двукратно. В точках, в которых сливаются два нуля функции $\wp(z) - c$, производная $\wp'(z)$ обращается в нуль. Фундаментальный параллелограмм можно выбрать так, чтобы внутри него лежали ровно четыре точки, для которых $u \equiv -u \pmod{\Lambda}$, а именно точки

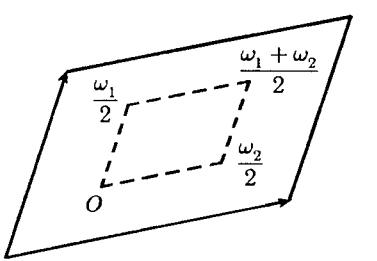


Рис. 20

$$0, \quad \frac{\omega_1}{2}, \quad \frac{\omega_2}{2} \quad \text{и} \quad \frac{\omega_1+\omega_2}{2}$$

(рис. 20). Первая из этих точек — полюс функции $\wp$, а три дру-

гие — нули функции $\wp'$. Итак, значения

$$e_1 = \wp\left(\frac{\omega_1}{2}\right), \quad e_2 = \wp\left(\frac{\omega_1+\omega_2}{2}\right) \quad \text{и} \quad e_3 = \wp\left(\frac{\omega_2}{2}\right)$$

для функции $\wp$ двукратные, причем других двукратных значений нет. Двукратные значения соответствуют нулям производной, поэтому $\wp'(z) = 0$ тогда и только тогда, когда

$$z \equiv \frac{\omega_1}{2}, \quad \frac{\omega_2}{2}, \quad \frac{\omega_1+\omega_2}{2} \pmod{\Lambda}.$$

Отметим, что числа e_1, e_2 и e_3 попарно различны. Предположим, например, что $e_1 = e_3$. Тогда функция $\wp(z) - e_1$ имеет двукратные нули в точках $\omega_1/2$ и $\omega_2/2$, т. е. внутри фундаментального параллелограмма расположено не менее четырех нулей этой функции, чего не может быть.

Функция Вейерштрасса не только дает пример эллиптической функции, но и позволяет описать, как устроены все эллиптические функции.

Теорема 3. *Пусть $f(z)$ — произвольная эллиптическая функция, $\wp(z)$ — функция Вейерштрасса с теми же периодами. Тогда существуют такие рациональные функции R и R_1, что*

$$f = R(\wp) + R_1(\wp)\,\wp'$$

Доказательство. Функцию $f(z)$ можно представить в виде суммы четной функции $g(z) = (f(z) + f(-z))/2$ и нечетной функции $h(z) = (f(z) - f(-z))/2$. А так как $\wp'(z)$ — нечетная функция, то $h_1(z) = h(z)/\wp'(z)$ — четная функция, причем

$$f(z) = g(z) + h_1(z)\,\wp'(z)$$

Поэтому достаточно доказать, что четную эллиптическую функцию можно представить в виде рациональной функции от $\wp$.

Докажем предварительно некоторые свойства нулей и полюсов четных эллиптических функций.

1. Пусть f — четная функция, u — ее нуль, соответственно полюс, порядка m. Тогда $-u$ тоже ее нуль, соответственно полюс порядка m. В случае нулей достаточно заметить, что для четной функции f справедливо равенство $f^{(k)}(-z) = (-1)^k f^{(k)}(z)$. В случае полюсов вместо f можно рассмотреть $1/f$.

2. Если f — четная эллиптическая функция и при этом $u \equiv -u \pmod{\Lambda}$, то порядок нуля или полюса функции f в точке u четен. Доказательство проведем для нулей (для полюсов вместо f можно рассмотреть $1/f$). Условие $u \equiv -u \pmod{\Lambda}$ эквивалентно тому, что

$$u \equiv 0, \quad \frac{\omega_1}{2}, \quad \frac{\omega_1+\omega_2}{2}, \quad \frac{\omega_2}{2} \pmod{\Lambda}.$$

Кроме того, из периодичности функции f' следует, что $f'(u) = f'(-u)$. Но производная четной функции нечетна, поэтому $f'(u) = 0$. Следовательно, если функция f имеет в точке u нуль, то он по крайней мере двукратный. При

$$u \equiv \frac{\omega_1}{2}, \quad \frac{\omega_1+\omega_2}{2}, \quad \frac{\omega_2}{2} \pmod{\Lambda}$$

функция $F(z) = \wp(z) - \wp(u)$ имеет в точке u нуль второго порядка; при $u \equiv 0 \pmod{\Lambda}$ таким свойством обладает функция $F(z) = 1/\wp(z)$. С помощью функции F можно построить четную эллиптическую функцию $f_1(z) = f(z)/F(z)$, порядок нуля в точке u для которой ровно на 2 меньше, чем для функции f. Поэтому если $f_1(u) \neq 0$, то порядок нуля функции f в точке u равен 2, а если $f_1(u) = 0$, то к функции f_1 можно применить такие же рассуждения, как и к функции f, и т. д.

Согласно доказанным выше свойствам нулей и полюсов четной эллиптической функции f их можно разбить на пары чисел, отличающихся друг от друга знаком, т. е. $(x, -x)$. Выберем в каждой такой паре по одному представителю. Пусть $a_1, \ldots, a_k$ — представители нулей, $b_1, \ldots, b_k$ — представители полюсов. Рассмотрим эллиптическую функцию

$$Q(z) = R(\wp(z)) = \frac{\prod(\wp(z) - \wp(a_i))}{\prod(\wp(z) - \wp(b_i))},$$

где в произведения входят лишь элементы a_i и b_i, отличные от узлов решетки (в узлах решетки функция $\wp$ принимает бесконечные значения). Если не обращать внимания на узлы решетки, то полная система нулей и полюсов у функции Q такая же, как и у функции f, поскольку $\wp(z) = \wp(a)$ тогда и только тогда, когда $z \equiv \pm a \pmod{\Lambda}$. Но согласно теореме 2, б) параграфа 2 сумма порядков нулей и полюсов эллиптической функции, лежащих внутри фундаментального параллелограмма, равна нулю,

поэтому порядок нуля или полюса в узле решетки однозначно определен порядками остальных нулей и полюсов. Следовательно, $f(z)/Q(z)$ — эллиптическая функция, не имеющая полюсов, т. е. константа. В итоге получаем, что $f(z) = cR(\wp(z))$. □

Задачи

Эллиптическая функция f такова, что все ее полюсы расположены в узлах решетки периодов. Докажите, что $f = P(\wp) + P_1(\wp)\wp'$, где P и P_1 — многочлены.

Указание. Из разложения $f = R(\wp) + R_1(\wp)\,\wp'$ можно получить, что $2R(\wp(z)) = f(z) + f(-z)$ и $2\wp'(z)R_1(\wp(z)) = f(z) - f(-z)$. Покажите, что функции R и R_1 не принимают бесконечных значений в конечных точках.

§ 4. Дифференциальное уравнение для функции $\wp(z)$

В предыдущем параграфе было доказано, что четная эллиптическая функция рационально выражается через $\wp(z)$, причем выражение было указано в явном виде. Это утверждение можно применить к четной функции $(\wp'(z))^2$. Она имеет двукратные нули в точках $\frac{\omega_1}{2}$, $\frac{\omega_1+\omega_2}{2}$, $\frac{\omega_2}{2}$ и шестикратный полюс в узле решетки. Следовательно,

$$(\wp'(z))^2 = c(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3), \qquad (4.1)$$

где

$$e_1 = \wp\left(\frac{\omega_1}{2}\right), \quad e_2 = \wp\left(\frac{\omega_1+\omega_2}{2}\right), \quad e_3 = \wp\left(\frac{\omega_2}{2}\right).$$

А так как

$$\wp(z) = z^{-2} + \dots$$

и

$$\wp'(z) = -2z^{-3} + \dots,$$

то $c = 4$.

Дифференциальное уравнение для функции $\wp(z)$ можно получить и другим способом. Заодно мы получим другую форму этого уравнения. Воспользуемся тем, что если коэффициенты при неположительных степенях z в разложениях Лорана функций

$(\wp'(z))^2$ и $a\wp^3(z) + b\wp^2(z) + c\wp(z) + d$ совпадают, то эти функции равны. В самом деле, их разность — эллиптическая функция без полюсов, в нуле принимающая нулевое значение. Следовательно, их разность — константа, причем нулевая.

Так как

$$\left(\frac{1}{1-x}\right)^2 = \frac{d}{dx}\left(\frac{1}{1-x}\right) = 1 + 2x + 3x^2 + \dots,$$

то

$$\wp(z) = \frac{1}{z^2} + {\sum}'\left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2}\right) =$$

$$= \frac{1}{z^2} + {\sum}'\left(\frac{1}{\omega^2}\left(1 + 2\frac{z}{\omega} + 3\left(\frac{z}{\omega}\right)^2 + \dots\right) - \frac{1}{\omega^2}\right) =$$

$$= \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + \dots,$$

где $G_k = {\sum}'\omega^{-k}$ (для нечетного k эта сумма нулевая). Поэтому

$$\wp(z) = z^{-2} + \dots,$$
$$\wp^2(z) = z^{-4} + 6G_4 + \dots,$$
$$\wp^3(z) = z^{-6} + 9G_4 z^{-2} + 15G_6 + \dots,$$
$$(\wp'(z))^2 = 4z^{-6} - 24G_4 z^{-2} - 80G_6 + \dots$$

(записаны лишь интересующие нас члены разложения Лорана). Таким образом,

$$a\wp^3(z) + b\wp^2(z) + c\wp(z) + d =$$
$$= az^{-6} + bz^{-4} + (9aG_4 + c)z^{-2} + (15aG_6 + 6bG_4 + d) + \dots$$

Следовательно, $a\wp^3 + b\wp^2 + c\wp + d = (\wp')^2$, если

$$\begin{cases} a = 4, \\ b = 0, \\ 9aG_4 + c = -24G_4, \\ 15aG_6 + 6bG_4 + d = -80G_6. \end{cases}$$

Полученная система уравнений, очевидно, имеет решение

$$\begin{cases} a = 4, \\ b = 0, \\ c = -60G_4, \\ d = -140G_6. \end{cases}$$

Для упрощения обозначений, обычно полагают

$$g_2 = 60G_4 = 60\sum{}'\omega^{-4},$$

$$g_3 = 140G_6 = 140\sum{}'\omega^{-6}.$$

Тогда

$$(\wp'(z))^2 = 4\wp^3(z) - g_2\wp(z) - g_3. \tag{4.2}$$

Сравнение выражений (4.1) и (4.2) показывает, что

$$e_1 + e_2 + e_3 = 0, \quad e_1e_2 + e_2e_3 + e_3e_1 = -g_2/4, \quad e_1e_2e_3 = g_3/4.$$

Легко проверить, что

$$g_2^3 - 27g_3^2 = 16(e_1 - e_2)^2(e_2 - e_3)^2(e_3 - e_1)^2.$$

В предыдущем параграфе было показано, что числа e_1, e_2 и e_3 попарно различны. Поэтому $g_2^3 - 27g_3^2 \neq 0$. Возникает естественный вопрос: если g_2 и g_3 — данные числа, причем $g_2^3 \neq 27g_3^2$, то обязательно ли существует решетка, для которой $g_2 = 60\sum'\omega^{-4}$ и $g_3 = 140\sum'\omega^{-6}$? Ответ на этот вопрос положительный (см., например, [Б15]).

Задачи

1. Докажите, что $\wp'' = 6\wp^2 - g_2/2$ и $\wp''' = 12\wp'\wp$.

2. Докажите, что $\wp^{(2n-2)}(z)$ и $\wp^{(2n+1)}(z)/\wp'(z)$ — многочлены степени n от $\wp(z)$.

3. Докажите, что $e_1^2 + e_2^2 + e_3^2 = g_2/2$, $e_1^3 + e_2^3 + e_3^3 = 3g_3/4$ и $e_1^4 + e_2^4 + e_3^4 = g_2^2/8$.

§ 5. Параметризация кубической кривой с помощью функции Вейерштрасса

Дифференциальное уравнение для функции $\wp$ позволяет прояснить природу сложения точек кубической кривой. Но для этого придется воспользоваться тем фактом, который мы оставили без доказательства: для любых чисел g_2 и g_3, где $g_2^3 \neq 27g_3^2$,

существует решетка, для которой функция Вейерштрасса удовлетворяет уравнению

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3.$$

Кубическую кривую

$$y^2 = 4x^3 - g_2x - g_3$$

можно параметризовать с помощью функции $\wp$, положив

$$x = \wp(z), \qquad y = \wp'(z).$$

Переходя к однородным координатам в $\mathbb{C}P^2$, отображение $f\colon \mathbb{C}/\Lambda \longrightarrow \mathbb{C}P^2$ определим следующим образом

$$z \mapsto (\wp(z), \wp'(z), 1) \quad \text{при} \quad z \neq 0,$$
$$z \mapsto (0, 1, 0) \quad \text{при} \quad z = 0.$$

Очевидно, что это отображение аналитично во всех точках, отличных от узлов решетки. Записав его в виде

$$z \mapsto \left(\frac{\wp(z)}{\wp'(z)},\ 1,\ \frac{1}{\wp'(z)}\right),$$

можно убедиться, что оно аналитично и в окрестности узла решетки. Отображение f взаимно однозначно отображает тор $\mathbb{C}/\Lambda$ на кубическую кривую

$$y^2z = 4x^3 - g_2xz^2 - g_3z^3$$

в $\mathbb{C}P^2$. В самом деле, на бесконечно удаленной прямой $z = 0$ лежит лишь точка $(0, 1, 0)$ этой кривой. В нее отображаются узлы решетки, которым на торе соответствует одна точка. Для остальных точек можно рассмотреть аффинную кривую

$$y^2 = 4x^3 - g_2x - g_3$$

и отображение $z \mapsto (\wp(z), \wp'(z))$. Уравнение $\wp(z) = c$ может иметь одно или два решения. Два решения оно имеет в том случае, когда $\wp'(z) \neq 0$. Решения при этом имеют вид $\pm z$. Образы этих двух точек при отображении $z \mapsto (\wp(z), \wp'(z))$ не совпадают, так как ненулевые числа $\wp'(z)$ и $\wp'(-z) = -\wp'(z)$ отличаются знаком.

Сложение точек комплексной плоскости индуцирует сложение точек тора, а оно, в свою очередь, с помощью отображения f индуцирует сложение точек кубической кривой. Оказывается, что это и есть определенное ранее сложение точек кубической кривой, причем в качестве нулевого элемента берется бесконечно удаленная точка $(0,1,0)$. Действительно, пусть точки P_1 и P_2 кубической кривой соответствуют точкам z_1 и z_2 комплексной плоскости, т. е. $P_i = (\wp(z_i), \wp'(z_i))$. Проведем через них прямую $y = ax + b$, при этом $\wp'(z_i) = a\wp(z_i) + b$, где $i = 1, 2$. Эллиптическая функция $\wp'(z) - a\wp(z) - b$ имеет в точке $z = 0$ полюс кратности 3, а в других точках фундаментального параллелограмма полюсов у нее нет. Следовательно, ранг этой функции равен 3, т. е. она имеет ровно 3 нуля, а именно, уже известные нам нули z_1 и z_2 и некоторый третий нуль z_3. А так как сумма полюсов равна нулю, то

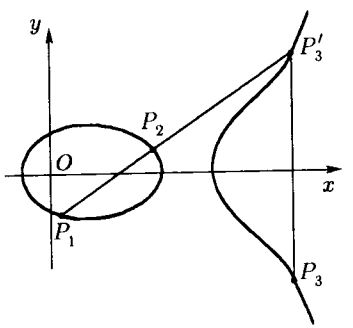


Рис. 21

$$z_1 + z_2 + z_3 \equiv 0 \pmod{\Lambda},$$

а значит, $z_3 \equiv -z_1 - z_2 \pmod{\Lambda}$. Итак, третьей точкой пересечения прямой P_1P_2 с кубической кривой будет точка

$$P_3' = (\wp(z_3), \wp'(z_3)) = (\wp(-z_1 - z_2), \wp'(-z_1 - z_2)) = (\wp(z_1 + z_2), -\wp'(z_1 + z_2)).$$

Таким образом, точка $P_3 = (\wp(z_1 + z_2), \wp'(z_1 + z_2))$, соответствующая сумме чисел z_1 и z_2, симметрична точке P_3' относительно оси Ox (рис. 21). Иными словами, P_3 — точка пересечения кубической кривой с прямой $P_3'E$, где $E = (0,1,0)$ — бесконечно удаленная точка кубической кривой. Именно в этом мы хотели убедиться.

Продолжив предыдущие рассуждения чуть дальше, можно доказать, что функция $\wp$ обладает алгебраической теоремой сложения, т. е. $\wp(z_1 + z_2)$ алгебраически выражается через $\wp(z_1)$ и $\wp(z_2)$. В самом деле, прямая $y = ax + b$, проходящая через

точки P_1 и P_2, пересекает кубическую кривую

$$y^2 = 4x^3 - g_2x - g_3$$

в трех точках (x_i, y_i), причем $x_1 = \wp(z_1)$, $x_2 = \wp(z_2)$ и $x_3 = \wp(z_1 + z_2)$. Поэтому кубическое уравнение

$$(ax + b)^2 = 4x^3 - g_2x - g_3$$

имеет указанные корни x_1, x_2 и x_3. Выразив через них коэффициент при x^2, получим

$$\wp(z_1) + \wp(z_2) + \wp(z_1 + z_2) = a^2/4.$$

А так как $\wp'(z_1) = a\wp(z_1) + b$ и $\wp'(z_2) = a\wp(z_2) + b$, то $a = \dfrac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)}$. Следовательно,

$$\wp(z_1 + z_2) = -\wp(z_1) - \wp(z_2) + \frac{1}{4}\left(\frac{\wp'(z_1) - \wp'(z_2)}{\wp(z_1) - \wp(z_2)}\right)^2.$$

Таким образом, $\wp(z_1 + z_2)$ рационально выражается через $\wp(z_i)$ и $\wp'(z_i)$, $i = 1, 2$. Остается напомнить, что $\wp'(z_i)$ алгебраически выражается через $\wp(z_i)$, а именно:

$$\wp'(z_i) = \sqrt{4\wp^3(z_i) - g_2\wp(z_i) - g_3}.$$

Замечание. С помощью функции Вейерштрасса можно параметризовать также и кривую $y^2 = G_4(x)$, где

$$G_4(x) = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4$$

— многочлен четвертой степени без кратных корней. Сделаем для этого замену переменных $x = x_1^{-1} + \alpha$, $y = y_1x_1^{-2}$. В результате получим

$$y_1^2x_1^{-4} = b_4x_1^{-4} + b_3x_1^{-3} + b_2x_1^{-2} + b_1x_1^{-1} + G_4(\alpha).$$

Если α — корень многочлена G_4, то

$$y_1^2 = b_1x_1^3 + b_2x_1^2 + b_3x_1 + b_4.$$

Эту кубическую кривую можно параметризовать с помощью функции Вейерштрасса.

Отметим, что аналогичным образом замена $x = x_1^{-1} + \alpha$, $y = y_1x_1^{-n}$ позволяет перейти от кривой $y^2 = G_{2n}(x)$ к кривой $y^2 = G_{2n-1}(x)$.

§ 6. Эллиптические интегралы

Функция Вейерштрасса $\wp(z)$ удовлетворяет дифференциальному уравнению

$$\left(\frac{d\wp}{dz}\right)^2 = 4\wp^3 - g_2\wp - g_3.$$

Следовательно,

$$dz = \frac{d\wp}{\sqrt{4\wp^3 - g_2\wp - g_3}},$$

т. е.

$$z = \int \frac{du}{\sqrt{4u^3 - g_2u - g_3}},$$

где $u = \wp(z)$. Таким образом, $z = \wp^{-1}(u)$, т. е. функция Вейерштрасса возникает при обращении интеграла

$$\int \frac{du}{\sqrt{4u^3 - g_2u - g_3}}.$$

Эллиптическим интегралом называют интеграл вида

$$\int R\left(x, \sqrt{G(x)}\right) dx,$$

где $G(x)$ — многочлен третьей или четвертой степени, не имеющий кратных корней, а $R(x, y)$ — рациональная функция двух переменных. Впервые такие интегралы появились при вычислении дуг различных кривых, в частности, эллипсов. Лишь позднее было замечено, что для некоторых эллиптических интегралов обратные функции обладают многими интересными свойствами, прежде всего, двоякопериодичностью.

Эллиптические интегралы можно привести к некоторым более простым формам. Но предварительно рассмотрим интегралы более простого вида. Прежде всего докажем, что если $R(x)$ — рациональная функция, то $\int R(x)\,dx$ есть сумма рациональной функции и некоторого числа слагаемых вида $c_i \ln(x - a_i)$. Для этого достаточно доказать, что рациональную функцию $R(x)$ можно представить в виде

$$A(x) + \sum_{i,k} \frac{c_{ik}}{(x - a_i)^k},$$

где $A(x)$ — многочлен.

Пусть $R(x) = P(x)/Q(x)$, где P и Q — многочлены. Поделив P на Q с остатком, можно перейти к дроби S/Q, где $\deg S < \deg Q$. Пусть $Q = Q_1Q_2$, где Q_1 и Q_2 — взаимно простые многочлены. Тогда существуют такие многочлены a и b, что $a(x)\,Q_1(x) + b(x)\,Q_2(x) = 1$. Поэтому

$$\frac{S}{Q_1Q_2} = \frac{a_1SQ_1 + a_2SQ_2}{Q_1Q_2} = \frac{a_1S}{Q_2} + \frac{a_2S}{Q_1}.$$

В полученных дробях нужно снова поделить с остатком числитель на знаменатель. После нескольких таких операций придем к сумме многочлена $A(x)$ и нескольких дробей вида $p(x)\,(x-a)^{-n}$, где $\deg p(x) < n$. Остается записать многочлен $p(x)$ в виде

$$p(x) = b_1(x - a)^{n-1} + b_2(x - a)^{n-2} + \ldots + b_n.$$

Первым проблемой интегрирования рациональных функций начал заниматься Лейбниц. Он рассматривал лишь разложения многочленов на сомножители с вещественными коэффициентами, поэтому у него возник вопрос: любой ли вещественный многочлен можно разложить на сомножители первой и второй степени с вещественными коэффициентами. В 1702 г. Лейбниц опубликовал статью, в которой утверждал, что многочлен $x^4 + a^4$ нельзя разложить требуемым образом, поскольку

$$x^4 + a^4 = (x^2 + a^2\sqrt{-1})\,(x^2 - a^2\sqrt{-1}) = \left(x + a\sqrt{\sqrt{-1}}\right) \times$$
$$\times \left(x - a\sqrt{\sqrt{-1}}\right)\left(x + a\sqrt{-\sqrt{-1}}\right)\left(x - a\sqrt{-\sqrt{-1}}\right),$$

а произведение никаких двух из этих множителей, как он считал, не может быть квадратным трехчленом с вещественными коэффициентами. И лишь через 17 лет Николай Бернулли указал, что

$$x^4 + a^4 = (x^2 + a^2)^2 - 2a^2x^2 = (x^2 + \sqrt{2}ax + a^2)\,(x^2 - \sqrt{2}ax + a^2).$$

В переписке Лейбница и Якоба Бернулли обсуждались также интегралы иррациональных выражений, возникающие при решении различных физических и математических задач. Многие из этих интегралов являются эллиптическими.

Перейдем теперь от рациональных функций к простейшим иррациональностям. Для вычисления интеграла

$$\int R\left(x, \sqrt{G(x)}\right)\, dx,$$

где $G(x) = ax + b$ — линейная функция, сделаем сначала замену $u = ax + b$. В результате получим интеграл вида

$$\int R_1\left(u, \sqrt{u}\right)\, du,$$

где R_1 снова рациональная функция. Сделаем затем замену $t = \sqrt{u}$. Тогда $du = d(t^2) = 2t\,dt$, а значит,

$$\int R_1\left(u, \sqrt{u}\right)\, du = \int R_1(t^2, t)\, 2t\, dt = \int R_2(t)\, dt,$$

где R_2 — рациональная функция.

Пусть теперь $G(x) = ax^2 + bx + c$. Как было сказано в предыдущем параграфе, с помощью замены

$$x = x_1^{-1} + \alpha, \quad y = y_1 x_1^{-1}$$

от кривой $y^2 = G(x)$ можно перейти к кривой $y_1^2 = G_1(x_1)$, где G_1 — линейная функция. Применим эту замену для вычисления интеграла $\int R(x, y)\, dx$, где $y^2 = G(x)$. Пусть

$$x = x_1^{-1} + \alpha, \quad y = y_1 x_1^{-1},$$

причем $G(\alpha) = 0$. Тогда $dx = -x_1^{-2} dx_1$ и

$$\int R(x, y)\, dx = -\int R(x_1^{-1} + \alpha, y_1 x_1^{-1}) x_1^{-2}\, dx_1 = \int R_1(x_1, y_1)\, dx_1,$$

причем $y_1^2 = Ax_1 + B$.

Итак, интегралы вида $\int R(x, y)\, dx$, где R — рациональная функция и $y = \sqrt{G(x)}$, выражаются через элементарные функции, если $\deg G \leqslant 2$. В случае, когда $\deg G = 3$, могут возникать функции, обратные к эллиптическим. Интеграл $\int R(x, y)\, dx$,

где $y = \sqrt{G_4(x)}$, сводится к интегралу $\int Q(x, y)\, dx$, где $y = \sqrt{4x^3 - g_2 x - g_3}$. В самом деле, с помощью замены $x = x_1^{-1} + \alpha$, $y = y_1 x_1^{-2}$ от многочлена G_4 четвертой степени можно перейти к многочлену третьей степени, а от произвольного многочлена третьей степени с помощью линейной замены можно перейти к многочлену вида $4x^3 - g_2 x - g_3$.

Можно было бы ограничиться вычислением интегралов вида $\int R(x, y)\, dx$, где $y^2 = 4x^3 - g_2 x - g_3$, но во многих случаях бывают полезны другие формы эллиптических интегралов. Поэтому мы сначала проведем вычисление эллиптических интегралов в общем виде, и лишь затем обратимся к некоторым их конкретным формам.

Пусть $I = \int R(x, y)\, dx$, где R — рациональная функция и

$$y^2 = a_0 x^4 + 4a_1 x^3 + 6a_2 x^2 + 4a_3 x + a_4,$$

причем хотя бы один из коэффициентов a_0 и a_1 не равен нулю.

Теорема 1 (Лежандр). *Эллиптический интеграл I можно представить в виде линейной комбинации рациональной функции от x и y, интеграла рациональной функции от x и интегралов*

$$\int \frac{dx}{y}, \quad \int \frac{x\, dx}{y}, \quad \int \frac{x^2\, dx}{y} \quad \text{и} \quad \int \frac{dx}{(x - c)\, y}.$$

Доказательство. Так как y^2 полиномиально выражается через x, то можно считать, что рациональная функция R не содержит y^k при $k \geqslant 2$. Кроме того,

$$\frac{a + by}{c + dy} = \frac{(a + by)(c - dy)\, y}{(c + dy)(c - dy)\, y} = \frac{A}{y} + B,$$

где A и B — рациональные функции от x. Поэтому вычисление исходного интеграла $\int R(x, y)\, dx$ сводится к вычислению интегралов вида $\int B(x)\, dx$ и $\int \frac{A(x)\, dx}{y}$. Рациональную функцию $A(x)$ можно представить в виде

$$A(x) = \sum a_n x^n + \sum \frac{a_{rm}}{(x - c_r)^m}.$$

Поэтому остается рассмотреть интегралы

$$J_n = \int \frac{x^n\,dx}{y}, \quad n \geqslant 0,$$

и

$$H_m = \int \frac{dx}{(x-c)^m y}, \quad m \geqslant 1.$$

Так как

$$\frac{d}{dx}(x^m y) = mx^{m-1}y + x^m \frac{dy}{dx} =$$

$$= \frac{1}{y}\left[mx^{m-1}y^2 + \frac{1}{2}x^m \frac{d(y^2)}{dx}\right] =$$

$$= a_0(m+2)\frac{x^{m+3}}{y} + 2a_1(2m+3)\frac{x^{m+2}}{y} +$$

$$+ 6a_2(m+1)\frac{x^{m+1}}{y} + 2a_3(2m+1)\frac{x^m}{y} + ma_4\frac{x^{m-1}}{y},$$

то получаем, что

$$x^m y = a_0(m+2)J_{m+3} + 2a_1(2m+3)J_{m+2} +$$

$$+ 6a_2(m+1)J_{m+1} + 2a_3(2m+1)J_m + ma_4 J_{m-1}.$$

Записав такие равенства для $m = 0, 1, 2, \ldots$, мы сможем последовательно выразить J_3 через J_0, J_1, J_2 и рациональную функцию от x и y, затем J_4 через J_0, J_1, J_2 и т. д. В случае, когда $a_0 = 0$, мы выразим J_2 через J_0 и J_1, затем J_3 через J_0 и J_1 и т. д.

Для вычисления интегралов $H_m = \int \frac{dx}{(x-c)^m y}$ запишем многочлен $G(x)$ в виде

$$G(x) = b_0(x-c)^4 + 4b_1(x-c)^3 + 6b_2(x-c)^2 + 4b_3(x-c) + b_4$$

(при этом $b_0 = a_0$). Как и в предыдущем случае, получим тождество

$$\frac{d}{dx}\left((x-c)^m y\right) = b_0(m+2)\frac{(x-c)^{m+3}}{y} +$$

$$+ 2b_1(2m+3)\frac{(x-c)^{m+2}}{y} + 6b_2(m+1)\frac{(x-c)^{m+1}}{y} +$$

$$+ 2b_3(2m+1)\frac{(x-c)^m}{y} + mb_4\frac{(x-c)^{m-1}}{y}.$$

Интегрируя эти тождества при $m = -1, -2, -3, \ldots$, получаем

$$\frac{y}{x-c} = b_0\int\frac{(x-c)^2}{y}dx + 2b_1\int\frac{x-c}{y}dx - 2b_3H_1 - b_4H_2,$$

$$\frac{y}{(x-c)^2} = 2b_1J_0 - 6b_2H_1 - 6b_3H_2 - 2b_4H_3,$$

$$\frac{y}{(x-c)^3} = -b_0J_0 + 6b_1cJ_0 - 6b_1J_1 - 12b_2H_2 - 10b_3H_3 - 3b_4H_4$$

и т. д. Эти равенства позволяют выразить $H_2, H_3, H_4, \ldots$, через J_0, J_1, $J_2, \ldots$, H_1 и рациональные функции x и y. □

Как мы уже говорили, любой эллиптический интеграл можно свести к интегралу $\int R(x,y)\,dx$, где $y^2 = 4x^3 - g_2x - g_3$. Эту форму эллиптических интегралов называют *формой Вейерштрасса*. Так как в этом случае $a_0 = 0$, то J_2 выражается через J_0 и J_1, поэтому остаются три вида интегралов

$$\int\frac{dx}{\sqrt{4x^3 - g_2x - g_3}},$$

$$\int\frac{x\,dx}{\sqrt{4x^3 - g_2x - g_3}},$$

$$\int\frac{dx}{(x-c)\sqrt{4x^3 - g_2x - g_3}}.$$

Другая широко распространенная форма эллиптических интегралов — это *форма Лежандра*. Для нее

$$y^2 = (1-x^2)(1-k^2x^2).$$

Перейти от формы Вейерштрасса к форме Лежандра можно следующим образом. С помощью замены $x = ax_1 + b$ от многочлена $4x^3 - g_2x - g_3$ перейдем к многочлену $x_1(x_1-1)(x_1-k^2)$. Затем сделаем замену $\xi^2 = x_1^{-1}$, $\eta^2 = y^2x_1^{-3}$. В результате получим $\eta^2 = (1-\xi^2)(1-k^2\xi^2)$.

Для формы Лежандра появляются все четыре вида интегралов:

$$\int\frac{dx}{\sqrt{G(x)}}, \quad \int\frac{x\,dx}{\sqrt{G(x)}}, \quad \int\frac{x^2dx}{\sqrt{G(x)}}, \quad \int\frac{dx}{(x-c)\sqrt{G(x)}},$$

где $G(x) = (1-x^2)(1-k^2x^2)$. Но при этом

$$\int \frac{x\,dx}{\sqrt{(1-x^2)(1-k^2x^2)}} = \frac{1}{2}\int \frac{du}{\sqrt{(1-u)(1-k^2u)}},$$

где $u = x^2$. Следовательно, этот интеграл выражается через элементарные функции.

Чтобы несколько упростить вид интегралов в форме Лежандра, сделаем замену $x = \sin\varphi$. Тогда $dx = \cos\varphi\,d\varphi$, $\sqrt{1-x^2} = \cos\varphi$ и $\sqrt{1-k^2x^2} = \sqrt{1-k^2\sin^2\varphi}$. Поэтому вышеуказанные неэлементарные интегралы принимают вид

$$\int \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}, \tag{6.1}$$

$$\int \frac{\sin^2\varphi\,d\varphi}{\sqrt{1-k^2\sin^2\varphi}}, \tag{6.2}$$

$$\int \frac{d\varphi}{(\sin\varphi - c)\sqrt{1-k^2\sin^2\varphi}}. \tag{6.3}$$

Интегралы (6.1) и (6.3) называются соответственно эллиптическими интегралами *первого* и *третьего рода*. Что же касается интеграла (6.2), то его можно представить в следующем виде

$$\int \frac{\sin^2\varphi\,d\varphi}{\sqrt{1-k^2\sin^2\varphi}} = \frac{1}{k^2}\int \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} - \frac{1}{k^2}\int \sqrt{1-k^2\sin^2\varphi}\,d\varphi.$$

Интеграл

$$\int \sqrt{1-k^2\sin^2\varphi}\,d\varphi \tag{6.4}$$

называется эллиптическим интегралом *второго рода*.

Для эллиптических интегралов первого и второго рода используются обозначения Лежандра:

$$F(\varphi) = \int_0^\varphi \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}, \qquad E(\varphi) = \int_0^\varphi \sqrt{1-k^2\sin^2\varphi}\,d\varphi.$$

Теорема 2 (следствие из теоремы Лежандра). *Всякий эллиптический интеграл I можно представить в виде линейной комбинации рациональной функции от x и y, интеграла рациональной функции от x и интегралов* (6.1), (6.2) *и* (6.4).

Задачи

1. Покажите, что интеграл $\int (1+x^6)^{-1/3}dx$ заменой переменных сводится к эллиптическому интегралу.

Указание. Сделайте замену $x^{-3} + x^3 = 2t^{-3/2}$.

2. Покажите, что интеграл $\int (1-x^3)^{-2/3}dx$ заменой переменных сводится к эллиптическому интегралу.

Указание. Сделайте замену $t(1-x) = (1-x^3)^{1/3}$.

§ 7. Теоремы сложения для эллиптических интегралов $F(\varphi)$ и $E(\varphi)$

Пусть

$$F(\varphi) = \int_0^\varphi \frac{d\varphi}{\Delta(\varphi)}, \qquad E(\varphi) = \int_0^\varphi \Delta(\varphi)\,d\varphi,$$

где

$$\Delta(\varphi) = \sqrt{1-k^2\sin^2\varphi}.$$

Теорема 1. *Пусть $F(\varphi) + F(\psi) = F(\mu)$. Тогда $\sin\mu$ алгебраически выражается через $\sin\varphi$ и $\sin\psi$.*

Доказательство. Рассмотрим дифференциальное уравнение

$$\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}} + \frac{d\psi}{\sqrt{1-k^2\sin^2\varphi}} = 0.$$

Оно имеет интеграл $F(\varphi) + F(\psi) - F(\mu) = 0$, где μ — константа. Учитывая нечетность функции F, его можно записать в виде $F(\varphi) + F(\psi) + F(-\mu) = 0$.

Покажем, что интегралом этого уравнения является также соотношение

$$\cos\varphi\cos\psi - \sin\varphi\sin\psi\sqrt{1-k^2\sin^2\varphi} = \cos\mu. \tag{7.1}$$

После возведения в квадрат оно становится симметричным по параметрам φ, ψ и $-\mu$:

$$\cos^2\varphi + \cos^2\psi + \cos^2\mu - 2\cos\varphi\cos\psi\cos\mu + k^2\sin^2\varphi\sin^2\psi\sin^2\mu = 1. \tag{7.2}$$

Следовательно, соотношение (7.1) выполняется вместе с соотношениями

$$\cos\mu\cos\varphi+\sin\mu\sin\varphi\sqrt{1-k^2\sin^2\psi}=\cos\psi, \tag{7.3}$$

$$\cos\mu\cos\psi+\sin\mu\sin\psi\sqrt{1-k^2\sin^2\varphi}=\cos\varphi. \tag{7.4}$$

Поделим обе части (7.1) на $\sin\varphi\sin\psi$ и продифференцируем полученное выражение. Результат можно преобразовать к виду

$$d\varphi\left(\frac{\cos\psi-\cos\mu\cos\varphi}{\sin\varphi}\right)+d\psi\left(\frac{\cos\varphi-\cos\mu\cos\psi}{\sin\psi}\right)=0.$$

Воспользовавшись формулами (7.3) и (7.4), получим

$$\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}+\frac{d\psi}{\sqrt{1-k^2\sin^2\psi}}=0.$$

Таким образом, выражение (7.1) действительно есть интеграл рассматриваемого дифференциального уравнения. Но двух независимых интегралов у него быть не может, поэтому из равенства $F(\varphi)+F(\psi)=F(\mu)$ следует равенство

$$\cos\varphi\cos\psi-\sin\varphi\sin\psi\sqrt{1-k^2\sin^2\mu}=\cos\mu.$$

Осталось заметить, что $\sin\mu$ алгебраически выражается через $\cos\mu$. □

В действительности, для $\sin\mu$ легко получить явное выражение. Пусть $x=\cos\mu$, тогда $\sin^2\mu=1-x^2$. Соотношение (7.2) можно рассматривать как квадратное уравнение относительно x. Решив его, получим

$$\cos\mu=\frac{\cos\varphi\cos\psi-\sin\varphi\sin\psi\,\Delta(\varphi)\,\Delta(\psi)}{1-k^2\sin^2\varphi\,\sin^2\psi}. \tag{7.5}$$

Корень квадратного уравнения выбран так, чтобы при малых φ и ψ формулы (7.5) и (7.1) были согласованы.

Несложными алгебраическими преобразованиями из (7.5) можно получить следующие выражения для $\sin\mu$ и $\Delta(\mu)$

$$\sin\mu=\frac{\sin\varphi\cos\psi\,\Delta(\psi)+\sin\psi\cos\varphi\,\Delta(\varphi)}{1-k^2\sin^2\varphi\,\sin^2\psi}, \tag{7.6}$$

$$\Delta(\mu)=\frac{\Delta(\varphi)\,\Delta(\psi)-k^2\sin\varphi\sin\psi\cos\varphi\cos\psi}{1-k^2\,\sin^2\varphi\sin^2\psi}. \tag{7.7}$$

Поделив (7.6) на (7.5), получим

$$\operatorname{tg}\mu=\frac{\operatorname{tg}\varphi\,\Delta(\varphi)+\operatorname{tg}\psi\,\Delta(\psi)}{1-\operatorname{tg}\varphi\operatorname{tg}\psi\,\Delta(\varphi)\,\Delta(\psi)}. \tag{7.8}$$

Последнюю формулу можно интерпретировать следующим образом. Пусть углы φ' и ψ' таковы, что $\operatorname{tg}\varphi'=\operatorname{tg}\varphi\,\Delta(\varphi)$ и $\operatorname{tg}\psi'=\operatorname{tg}\psi\,\Delta(\psi)$. Тогда $\mu=\varphi'+\psi'$.

В приложениях весьма важен случай $\mu=\pi/2$. В этом случае $\cos\mu=0$ и $\sin\mu=1$. Из (7.3) и (7.4) получаем $\sin\varphi=\cos\psi/\Delta(\psi)$ и $\sin\psi=\cos\varphi/\Delta(\varphi)$, а из (7.1) получаем $\cos\varphi\cos\psi=b\sin\varphi\sin\psi$, т. е. $b\operatorname{tg}\varphi\operatorname{tg}\psi=1$, где $b=\sqrt{1-k^2}$. Из (7.5) следует, что

$$\cos\varphi\cos\psi=\Delta(\varphi)\,\Delta(\psi)\,\sin\varphi\sin\psi,$$

поэтому $\Delta(\varphi)\,\Delta(\psi)=b$. Следовательно, $\cos\varphi=\sin\psi\Delta(\varphi)=b\sin\psi/\Delta(\psi)$ и $\cos\psi=b\sin\varphi/\Delta(\varphi)$.

Формулы (7.5) и (7.6) до некоторой степени напоминают формулы для косинуса и синуса суммы двух углов. С их помощью можно получить выражения, аналогичные выражениям $\cos n\varphi$ и $\sin n\varphi$ через $\cos\varphi$ и $\sin\varphi$. Пусть $F(\varphi_n)=nF(\varphi)$. Тогда

$$\sin\varphi_2=\frac{2\sin\varphi\cos\varphi\,\Delta(\varphi)}{1-k^2\sin^4\varphi},\quad \cos\varphi_2=\frac{1-2\sin^2\varphi+k^2\sin^4\varphi}{1-k^2\sin^4\varphi},$$

$$\operatorname{tg}\varphi_2=\frac{2\operatorname{tg}\varphi\,\Delta(\varphi)}{1-(\operatorname{tg}\varphi\,\Delta(\varphi))^2},\quad \Delta(\varphi_2)=\frac{1-2k^2\sin^2\varphi+k^2\sin^4\varphi}{1-k^2\sin^4\varphi}.$$

Для нахождения φ по данному φ_2 можно воспользоваться тем, что $\operatorname{tg}(\varphi_2/2)=\operatorname{tg}\varphi\,\Delta(\varphi)$, а можно также решить уравнение

$$\cos\varphi_2=\frac{1-2x^2+k^2x^4}{1-k^2x^4},$$

где $x=\sin\varphi$. Это уравнение соответствует делению $F(\varphi_2)$ пополам.

Для деления $F(\psi)$ на три равные части нужно решить уравнение

$$\sin\psi=\frac{3x-4(1+k^2)\,x^3+6k^2x^5-k^4x^9}{1-6k^2x^4+4k^2(1+k^2)\,x^6-3k^4x^8},$$

где $x = \sin\varphi$. При $\psi = \pi/2$ получим уравнение

$$(1+x)(1-2x+2k^2x^3-k^2x^4)^2 = 0,$$

т. е. деление $F(\pi/2)$ на три равные части сводится к решению уравнения

$$1 - 2\sin\varphi + 2k^2\sin^3\varphi + k^2\sin^4\varphi = 0.$$

Для эллиптического интеграла второго рода $E(\varphi)$ также есть теорема сложения, которая непосредственно связанная с теоремой сложения для эллиптических интегралов первого рода $F(\varphi)$.

Теорема 2. *Если* $F(\varphi) + F(\psi) - F(\mu) = 0$, *то*

$$E(\varphi) + E(\psi) - E(\mu) = k^2\sin\varphi\sin\psi\sin\mu.$$

Доказательство. Пусть $E(\varphi) + E(\psi) - E(\mu) = P(\varphi,\psi,\mu)$. Продифференцируем это равенство при постоянном значении μ. В результате получим

$$\Delta(\varphi)\,d\varphi + \Delta(\psi)\,d\psi = dP.$$

Но согласно (7.3), (7.4)

$$\Delta(\varphi) = \frac{\cos\varphi - \cos\psi\cos\mu}{\sin\psi\sin\mu}, \qquad \Delta(\psi) = \frac{\cos\psi - \cos\varphi\cos\mu}{\sin\varphi\sin\mu}.$$

Поэтому

$$dP = \left(\frac{\cos\varphi - \cos\psi\cos\mu}{\sin\psi\sin\mu}\right)d\varphi + \left(\frac{\cos\psi - \cos\varphi\cos\mu}{\sin\varphi\sin\mu}\right)d\psi = \frac{d(\sin^2\varphi + \sin^2\psi + 2\cos\varphi\cos\psi\cos\mu)}{2\sin\varphi\sin\psi\sin\mu}.$$

Согласно (7.2)

$$1 - \sin^2\varphi + 1 - \sin^2\psi - 2\cos\varphi\cos\psi\cos\mu = 1 - \cos^2\mu - k^2\sin^2\varphi\,\sin^2\psi\,\sin^2\mu.$$

Поэтому

$$dP = \frac{d(k\sin\varphi\sin\psi\sin\mu)^2}{2\sin\varphi\sin\psi\sin\mu} = k^2\,d(\sin\varphi\sin\psi\sin\mu).$$

А так как P и $\sin\varphi\sin\psi\sin\mu$ обращаются в нуль при $\varphi = 0$, то $P = k^2\sin\varphi\sin\psi\sin\mu$. □

Эта теорема сложения позволяет получить для разностей $nE(\varphi) - E(\varphi_n)$, где $F(\varphi_n) = nF(\varphi)$, следующие выражения:

$$2E(\varphi) - E(\varphi_2) = k^2\sin\varphi\sin\varphi\sin\varphi_2,$$

$$3E(\varphi) - E(\varphi_3) = (2E(\varphi) - E(\varphi_2)) + (E(\varphi_2) + E(\varphi) - E(\varphi_3)) = k^2\sin\varphi\,(\sin\varphi\sin\varphi_2 + \sin\varphi_2\sin\varphi_3),$$

и т. д.

Применениям полученных в этом параграфе формул для решения задач о делении дуг эллипса посвящена гл. 3.

§ 8. Эллиптические функции Якоби

В предыдущем параграфе мы рассмотрели функцию

$$F(\varphi) = \int_0^x \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}} = \int_0^\varphi \frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}}, \qquad (8.1)$$

где $x = \sin\varphi$, и получили для нее теорему сложения: если $F(\varphi) + F(\psi) = F(\mu)$, то

$$\sin\mu = \frac{\sin\varphi\cos\psi\,\Delta(\psi) + \sin\psi\cos\varphi\,\Delta(\varphi)}{1 - k^2\sin^2\varphi\sin^2\psi}. \qquad (8.2)$$

Частный вид формул (8.2) и (8.2) при $k = 0$ хорошо знаком:

$$\arcsin\varphi = \int_0^x \frac{dx}{\sqrt{1-x^2}} = \int_0^\varphi d\varphi,$$

где $x = \sin\varphi$, и

$$\sin\mu = \sin\varphi\cos\psi + \sin\psi\cos\varphi,$$

если $\mu = \varphi + \psi$.

Функция $x = \sin\varphi$ во многих отношениях лучше, чем $\arcsin x$. Неудивительно поэтому, что рассматривать функцию, обратную к $F(\varphi)$ удобнее, чем саму $F(\varphi)$. Заменим обозначение Лежандра $F(\varphi)$ на обозначение Якоби $u(\varphi) = F(\varphi)$. Обратная к $u(\varphi)$ функция $\varphi(u)$ называется *амплитудой* u и обозначается $\operatorname{am} u$. В предыдущем параграфе были получены формулы (7.5)–(7.7) для функций $\sin\varphi$, $\cos\varphi$ и $\Delta(\varphi) = \sqrt{1-k^2\sin^2\varphi}$. Если

ввести функции

$$\operatorname{sn} u = \sin \operatorname{am} u, \qquad \operatorname{cn} u = \cos \operatorname{am} u, \qquad \operatorname{dn} u = \Delta(\operatorname{am} u),$$

то эти формулы можно записать в следующем виде:

$$\operatorname{cn}(u+v) = \frac{\operatorname{cn} u \operatorname{cn} v - \operatorname{sn} u \operatorname{sn} v \operatorname{dn} u \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}, \tag{8.3}$$

$$\operatorname{sn}(u+v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}, \tag{8.4}$$

$$\operatorname{dn}(u+v) = \frac{\operatorname{dn} u \operatorname{dn} v - k^2 \operatorname{sn} u \operatorname{sn} v \operatorname{cn} u \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}. \tag{8.5}$$

Функции $\operatorname{sn} u$, $\operatorname{cn} u$ и $\operatorname{dn} u$ обычно называют *эллиптическими функциями Якоби*, хотя многие свойства этих функций в том или ином виде были обнаружены до него Лежандром и Абелем.

Одно из важнейших свойств функций $\operatorname{sn} u$, $\operatorname{cn} u$ и $\operatorname{dn} u$ — двоякопериодичность. Наличие у этих функций одного периода достаточно очевидно. В самом деле, функции $\sin\varphi$ и $\cos\varphi$ имеют период 2π, а функция $\sin^2\varphi$ имеет период π. Поэтому функции $\operatorname{sn} u$ и $\operatorname{cn} u$ имеют период $4K$, где

$$K = \int_0^{\pi/2} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2\varphi}},$$

а функция $\operatorname{dn} u = \sqrt{1 - k^2 \operatorname{sn}^2 u}$ имеет период $2K$.

С помощью теорем сложения (8.3)–(8.5) можно выяснить поведение функций $\operatorname{sn} u$, $\operatorname{cn} u$ и $\operatorname{dn} u$ при увеличении аргумента на четверть периода K и на полупериод $2K$. Подставив значения

$$\operatorname{sn} K = 1, \qquad \operatorname{cn} K = 0, \qquad \operatorname{dn} K = \sqrt{1 - k^2}$$

в формулы (8.3)–(8.5), получим

$$\operatorname{sn}(u+K) = \frac{\operatorname{cn} u}{\operatorname{dn} u},$$

$$\operatorname{cn}(u+K) = -\frac{\sqrt{1-k^2}\,\operatorname{sn} u}{\operatorname{dn} u},$$

$$\operatorname{dn}(u+K) = \frac{\sqrt{1-k^2}}{\operatorname{dn} u}.$$

А так как $\operatorname{sn} 2K = 0$, $\operatorname{cn} 2K = 1$ и $\operatorname{dn} 2K = 1$, то

$$\operatorname{sn}(u+2K) = -\operatorname{sn} u, \quad \operatorname{cn}(u+2K) = -\operatorname{cn} u, \quad \operatorname{dn}(u+2K) = \operatorname{dn} u.$$

Обнаружить второй период эллиптических функций Якоби несколько сложнее. Вспомним предварительно, что интеграл

$$\int \frac{d\varphi}{\sqrt{1 - k^2 \sin^2\varphi}}$$

был получен из интеграла

$$\int \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

заменой $x = \sin\varphi$. Сейчас нам будет удобнее работать с исходным интегралом. На комплексной плоскости функция

$$u(x) = \int_0^x \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

вообще говоря, не определена, потому что значение $u(x)$ зависит от пути интегрирования. Значения функции $u(x)$ в одной и той же точке могут отличаться на числа вида

$$L = \int_C \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}},$$

где интеграл берется по замкнутому контуру C. При этом любое такое число L будет периодом обратной функции $x(u)$. Подынтегральная функция имеет особые точки ± 1, $\pm k^{-1}$. Посмотрим, как влияют обходы вокруг этих точек на значения функций $\operatorname{sn} u = x$, $\operatorname{cn} u = \sqrt{1-x^2}$ и $\operatorname{dn} u = \sqrt{1-k^2x^2}$.

Пусть

$$K = \int_0^1 \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}$$

и

$$\alpha = \int_0^X \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

Обход вдоль пути, изображенного на рис. 22, показывает, что значение интеграла в точке X равно α и $K + (K - \alpha) = 2K - \alpha$. В самом деле, на последнем участке изменяется знак функции $\sqrt{1-x^2}$ и изменяется направление интегрирования, в результате двух изменений знак перед интегралом получается положительный. Таким образом, $\operatorname{sn}\alpha = \operatorname{sn}(2K - \alpha)$. Кроме того, знак функции $\sqrt{1-x^2}$ при этом обходе изменяется, а знак функции $\sqrt{1-k^2x^2}$ не изменяется. Поэтому $\operatorname{cn}\alpha = -\operatorname{cn}(2K-\alpha)$ и $\operatorname{dn}\alpha = \operatorname{dn}(2K-\alpha)$. Заменив α на $-\alpha$, получим

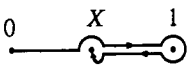


Рис. 22

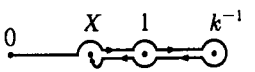


Рис. 23

$$\operatorname{sn}(\alpha+2K) = \operatorname{sn}(-\alpha) = -\operatorname{sn}\alpha,$$

$$\operatorname{cn}(\alpha+2K) = -\operatorname{cn}(-\alpha) = -\operatorname{cn}\alpha,$$

$$\operatorname{dn}(\alpha+2K) = \operatorname{dn}(-\alpha) = \operatorname{dn}\alpha.$$

Эти формулы мы уже получали другим способом.

Рассмотрим теперь обход вдоль пути, изображенного на рис. 23. Пусть

$$iK' = \int_1^{k^{-1}} \frac{dx}{\sqrt{(1-x^2)(1-k^2x^2)}}.$$

Число K' вещественно, так как при $x \in (1, k^{-1})$ число $\sqrt{1-x^2}$ мнимое. Тогда значение интеграла в точке X равны α и $K+iK'+iK'-(K-\alpha)$. Пояснений здесь требует лишь знак последнего слагаемого. Дело в том, что произошли три перемены знака: изменилось направление интегрирования и изменились знаки обеих функций $\sqrt{1-x^2}$ и $\sqrt{1-k^2x^2}$. В результате получаем

$$\operatorname{sn}\alpha = \operatorname{sn}(\alpha+2iK'),\ \operatorname{cn}\alpha = -\operatorname{cn}(\alpha+2iK'),\ \operatorname{dn}\alpha = -\operatorname{dn}(\alpha+2iK').$$

Следовательно, функции $\operatorname{sn}u$, $\operatorname{cn}u$ и $\operatorname{dn}u$ имеют периоды $2iK'$, $4iK'$ и $4iK'$ соответственно. Кроме того, у функции $\operatorname{cn}u$ есть период $2K+2iK'$. В самом деле,

$$\operatorname{cn}(\alpha+2iK'+2K) = -\operatorname{cn}(\alpha+2iK') = \operatorname{cn}\alpha.$$

Таким образом, функция $\operatorname{sn}u$ имеет периоды $4K$ и $2iK'$, функция $\operatorname{cn}u$ имеет периоды $4K$ и $2K+2iK'$, а функция $\operatorname{dn}u$ имеет периоды $2K$ и $4iK'$.

Задачи

Докажите, что $K' = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-k'^2t^2)}}$, где $k'^2+k^2=1$.

Указание. Сделайте замену $k't = \sqrt{1-k^2x^2}$, где число k' выбрано так, что $k\sqrt{x^2-1} = k'\sqrt{1-t^2}$.

§ 9. Теорема Вейерштрасса о функциях, обладающих алгебраической теоремой сложения

Будем говорить, что мероморфная функция $\varphi(z)$ *обладает алгебраической теоремой сложения*, если существует такой ненулевой многочлен F от трех переменных, что

$$F(\varphi(z_1+z_2), \varphi(z_1), \varphi(z_2)) = 0.$$

Это равенство означает, что $\varphi(z_1+z_2)$ алгебраически выражается через $\varphi(z_1)$ и $\varphi(z_2)$.

Например, алгебраической теоремой сложения обладает функция $\operatorname{sn}z$. В самом деле, если

$$a = \operatorname{sn}(z_1+z_2), \qquad b = \operatorname{sn}z_1, \qquad c = \operatorname{sn}z_2,$$

то

$$a = \frac{b\sqrt{1-c^2}\sqrt{1-k^2c^2} + c\sqrt{1-b^2}\sqrt{1-k^2b^2}}{1-k^2b^2c^2}.$$

С помощью двух возведений в квадрат можно избавиться от радикалов и получить полиномиальное соотношение $F(a,b,c)=0$.

Функция Вейерштрасса $\wp(z)$ тоже обладает алгебраической теоремой сложения. В самом деле, как показано в § 5, если

$$a = \wp(z_1+z_2), \quad b = \wp(z_1) \quad \text{и} \quad c = \wp(z_2),$$

то

$$a = -b-c+\frac{1}{4}\left(\frac{\sqrt{4b^3-g_2b-g_3}-\sqrt{4c^3-g_2c-g_3}}{b-c}\right)^2.$$

После преобразования этой формулы можно получить соотношение вида $F(a,b,c)=0$, где F — многочлен.

Любую эллиптическую функцию f можно записать в виде

$$f = R(\wp) + R_1(\wp)\,\wp',$$

где R и R_1 — рациональные функции (см. § 3). Это представление позволяет получить алгебраическую теорему сложения для произвольной эллиптической функции, если учесть, что

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3.$$

В самом деле, если $R_1 \neq 0$, то $\wp' = \dfrac{f - R(\wp)}{R_1(\wp)}$. Поэтому

$$\left(\frac{f - R(\wp)}{R_1(\wp)}\right)^2 = 4\wp^3 - g_2\wp - g_3.$$

Это соотношение можно переписать в виде $P(f,\wp)=0$, где P — многочлен, степень которого по f равна 2. В случае, когда $R_1 = 0$, соотношение $f = R(\wp)$ тоже можно записать в таком виде. Пусть

$$A = f(z_1+z_2), \quad B = f(z_1), \quad C = f(z_2).$$

Из соотношений $F(a,b,c)=0$ и $P(A,a)=0$ можно получить соотношение $F_1(A,b,c)=0$, вычислив результант многочленов $f(a)=F(a,b,c)$ и $g(a)=P(A,a)$. Затем из соотношений $F_1(A,b,c)=0$ и $P(B,b)=0$ получим, что $F_2(A,B,c)=0$, а из соотношений $F_2(A,B,c)=0$ и $P(C,c)=0$ получим требуемое соотношение $G(A,B,C)=0$.

Алгебраической теоремой сложения обладают не только эллиптические функции. Как хорошо известно, алгебраической теоремой сложения обладает функция e^z, так как $e^{z_1+z_2} = e^{z_1}e^{z_2}$. Кроме того, если $u = R(f)$, где R — рациональная функция, то $P(u,f)=0$, где P — многочлен линейный по u. Поэтому из соотношения $F(a,b,c)=0$ можно получить соотношение $G(A,B,C)=0$, где $A=R(a)$, $B=R(b)$ и $C=R(c)$. Поэтому любая рациональная функция, а также любая рациональная функция от $e^{\lambda z}$ обладает алгебраической теоремой сложения.

Оказывается, что приведенные примеры исчерпывают все мероморфные функции, обладающие алгебраической теоремой сложения.

Теорема 1 (Вейерштрасс). *Мероморфная функция $\varphi(z)$, обладающая алгебраической теоремой сложения, либо является эллиптической, либо имеет вид или $R(z)$, или $R(e^{\lambda z})$, где R — рациональная функция.*

Доказательство. В конечной области мероморфная функция не имеет особых точек, отличных от полюсов. Если существует

$$\lim_{z\to\infty}\varphi(z) \quad \text{или} \quad \lim_{z\to\infty}\frac{1}{\varphi(z)},$$

то функция $\varphi(z)$ рациональная. В самом деле, вычтем из функции φ сумму ее главных частей во всех полюсах (если точка ∞ является полюсом, то главная часть функции в этой точке имеет вид

$$a_r z^r + a_{r+1}z^{r+1} + \dots,$$

где $r>0$). В результате получим функцию f, не имеющую особых точек, причем точка ∞ тоже неособая. Следовательно, функция f — константа, а исходная функция φ — рациональная.

В дальнейшем будем считать, что функция φ не рациональная, т. е. точка ∞ существенно особая. Для доказательства теоремы Вейерштрасса нам потребуется следующая теорема о поведении функции в окрестности существенно особой точки.

Теорема 2 (Большая теорема Пикара). *Аналитическая функция $\varphi(z)$ принимает в произвольной окрестности существенно особой точки любое конечное значение, за исключением, быть может, одного.*

Доказательство этой теоремы можно найти в [Б3]. □

Пусть

$$F(\varphi(z_1+z_2),\varphi(z_1),\varphi(z_2)) = 0,$$

где F — многочлен, причем его степень по первому аргументу равна n. Нужно доказать, что функция φ периодическая, причем если она не двоякопериодическая, то $\varphi(z)=R(e^{\lambda z})$. Из теоремы Пикара следует, что функция φ в окрестности существенно особой точки принимает некоторое значение c бесконечно много раз. Пусть $a_1,\dots,a_{n+1}$ — точки, в которых φ принимает значение c. Особые точки функции φ, а также точки z, для которых точки $z+a_i$ особые, образуют множество меры нуль. Поэтому существует неособая точка z_0 функции φ, для которой точки a_i+z_0 тоже

неособые. При этом точки z и $a_i + z$ для z, достаточно близких к z_0, тоже неособые. Для таких точек z рассмотрим уравнение

$$F(x, \varphi(z), c) = 0. \tag{9.1}$$

Оно имеет $n+1$ корень $x_i = \varphi(z + a_i)$, так как

$$F(\varphi(z + a_i), \varphi(z), c) = F(\varphi(z + a_i), \varphi(z), \varphi(a_i)) = 0$$

Уравнение (9.1) представляет собой ненулевой многочлен степени n от x, поэтому оно имеет не более n различных корней. Следовательно, $\varphi(z + a_p) = \varphi(z + a_q)$ для некоторых несовпадающих p и q. Такие соотношения выполняются для любой точки z из окрестности точки z_0, но пары (p, q) могут при этом меняться. Тем не менее, пар (p, q) конечное число, поэтому некоторое соотношение $\varphi(z + a_p) = \varphi(z + a_q)$ выполняется для бесконечного набора точек z. Эти точки имеют предельную точку z_1, причем функция φ в ней регулярна. По теореме единственности получаем, что функции $\varphi(z + a_p)$ и $\varphi(z + a_q)$ совпадают, т. е. $a_p - a_q$ — период функции φ.

Мы доказали, что функция φ периодическая, для определенности можно считать, что минимальный период функции φ равен 2π. Предположим, что других периодов у функции φ нет. Докажем, что тогда $\varphi(z) = R(w)$, где $w = e^{iz}$, R — рациональная функция. Отображение $z \longmapsto w = e^{iz}$ переводит полосу $0 \leqslant \operatorname{Re} z < 2\pi$ в плоскость с разрезом от 0 до $+\infty$. Функция $\psi(w) = \varphi(z)$ мероморфна на плоскости с выколотыми точками 0 и ∞. Если эти точки не являются существенно особыми, то функция ψ рациональна. Предположим, что хотя бы одна из них существенно особая. Тогда согласно теореме Пикара существуют точки $b_1, \ldots, b_{n+1}$, для которых $\psi(b_i) = c$. Прообразы $\beta_1, \ldots, \beta_{n+1}$ этих точек при отображении $z \longmapsto w$ попарно различны и лежат в полосе $0 \leqslant \operatorname{Re} z < 2\pi$. Уравнение $F(x, \varphi(z), c) = 0$ имеет корни $x_i = \varphi(\beta_i + z)$. Повторив те же рассуждения, что и раньше, получим, что функция φ имеет период $\beta_p - \beta_q$, причем $0 \leqslant \operatorname{Re}\beta_p, \operatorname{Re}\beta_q < 2\pi$. Следовательно, функция φ либо имеет не только чисто вещественный период 2π, либо имеет вещественный период, меньший 2π. Последний случай невозможен, так как период 2π по предположению наименьший. □

ГЛАВА 3

ДУГИ КРИВЫХ И ЭЛЛИПТИЧЕСКИЕ ИНТЕГРАЛЫ

Для окружности легко построить дугу, длина которой равна сумме длин двух ее заданных дуг. Вообще говоря, это связано с тем, что $\sin(\varphi + \psi)$ выражается через $\sin\varphi$ и $\sin\psi$ по формуле

$$\sin(\varphi + \psi) = \sin\varphi\sqrt{1 - \sin^2\psi} + \sin\psi\sqrt{1 - \sin^2\varphi}.$$

Но, как мы уже знаем, для эллиптических интегралов $F(\varphi)$ и $E(\varphi)$ также есть теоремы сложения. Это означает, что для кривых, дуги которых выражаются через $F(\varphi)$ и $E(\varphi)$, также возможна операция сложения дуг, хотя она будет и более громоздкой. Для кривых, дуги которых выражаются через $E(\varphi)$, это будет сложение не в прямом смысле, а с некоторой алгебраической добавкой.

В этой главе мы займемся кривыми, дуги которых выражаются через эллиптические интегралы. Во многих отношениях из всех таких кривых наиболее интересна лемниската. Но она заслуживает специального изучения, и ей мы займемся в гл. 4.

§ 1. Дуги эллипса и гиперболы

Нетрудно проверить, что с помощью формул $x = a\cos\varphi$, $y = b\sin\varphi$ можно параметрически задать эллипс

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Дифференциал dl длины дуги эллипса равен

$$\sqrt{dx^2 + dy^2} = d\varphi\sqrt{a^2\cos^2\varphi + b^2\sin^2\varphi}.$$

Если $a = 1$ и $b = \sqrt{1 - k^2}$, то

$$dl = d\varphi\sqrt{1 - k^2\sin^2\varphi}.$$

В этом случае длина дуги эллипса, заключенной между концом малой полуоси B и точкой $M = (a\cos\varphi, b\sin\varphi)$, равна

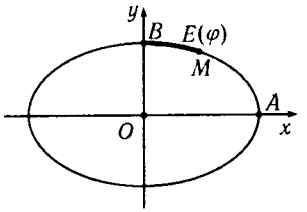


Рис. 24

$$E(\varphi) = \int_0^{\varphi} \sqrt{1 - k^2\sin^2\psi}\, d\psi.$$

Таким образом, длина дуги эллипса выражается эллиптическим интегралом второго рода. Именно отсюда пошло название для интегралов этого типа — эллиптический интеграл.

Наиболее простая параметризация гиперболы

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

получается с помощью гиперболических функций: $x = a\operatorname{ch} t$, $y = b\operatorname{sh} t$. Но для выражения длины гиперболы через $F(\varphi)$ и $E(\varphi)$ нужна параметризация с помощью тригонометрических функций. Одну из возможных таких параметризаций дают формулы $x = a/\cos\varphi$, $y = b\operatorname{tg}\varphi$. При этой параметризации дифференциал длины дуги равен

$$\frac{1}{\cos^2\varphi}\sqrt{a^2\sin^2\varphi + b^2}\, d\varphi,$$

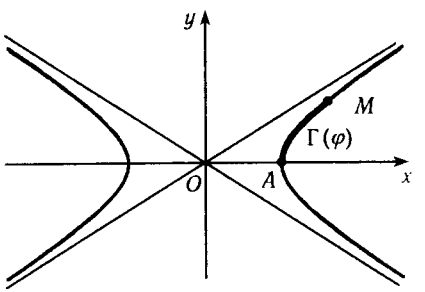


Рис. 25

желаемого выражения эта формула не дает. Рассмотрим поэтому другую параметризацию, положив $y = b^2\operatorname{tg}\varphi$. Тогда

$$x^2 = \left(\frac{a}{\cos\varphi}\right)^2 \left(1 - (1 - b^2)\sin^2\varphi\right).$$

В частности, в том случае, когда $a^2 = 1 - b^2 = k^2$, получим

$$x = \frac{k}{\cos\varphi}\sqrt{1 - k^2\sin^2\varphi} \quad \text{и} \quad y = (1 - k^2)\operatorname{tg}\varphi$$

а значит,

$$dl = \frac{(1 - k^2)\, d\varphi}{\cos^2\varphi\sqrt{1 - k^2\sin^2\varphi}}.$$

Поэтому длина дуги AM гиперболы (рис. 25) равна

$$\Gamma(\varphi) = \int_0^{\varphi} \frac{(1 - k^2)\, d\psi}{\cos^2\psi\sqrt{1 - k^2\sin^2\psi}} = \int_0^{\varphi} \frac{(1 - k^2)\, d\psi}{\cos^2\psi\,\Delta(\psi)},$$

где $\Delta(\psi) = \sqrt{1 - k^2\sin^2\psi}$. А так как $\Delta'(\psi) = -\dfrac{k^2\sin\psi\cos\psi}{\Delta(\psi)}$, то

$$(\Delta(\psi)\operatorname{tg}\psi)' = -\frac{k^2\sin^2\psi}{\Delta(\psi)} + \frac{\Delta(\psi)}{\cos^2\psi} = \frac{1 - k^2}{\cos^2\psi\,\Delta(\psi)} - \frac{1 - k^2}{\Delta(\psi)} + \Delta(\psi),$$

поэтому

$$\Gamma(\varphi) = \Delta(\varphi)\operatorname{tg}\varphi - \int_0^{\varphi} \Delta(\psi)\, d\psi + (1 - k^2)\int_0^{\varphi} \frac{d\psi}{\Delta(\psi)} =$$

$$= \Delta(\varphi)\operatorname{tg}\varphi - E(\varphi) + (1 - k^2)F(\varphi).$$

Таким образом, длина дуги гиперболы тоже выражается через эллиптические интегралы $E(\varphi)$, $F(\varphi)$ и элементарную функцию $\Delta(\varphi)\operatorname{tg}\varphi$, хотя и это выражение более сложное.

§ 2. Деление дуг эллипса

Рассмотрим эллипс $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} = 1$, где $a = 1$ и $b = \sqrt{1 - k^2}$.

Длина его дуги BM (рис. 24) равна $E(\varphi)$. В § 7 гл. 2 было показано, что если $F(\varphi) + F(\psi) = F(\mu)$, то

$$E(\varphi) + E(\psi) - E(\mu) = k^2\sin\varphi\sin\psi\sin\mu,$$

при этом φ, ψ и μ связаны соотношением

$$\cos\varphi\cos\psi - \sin\varphi\sin\psi\sqrt{1 - k^2\sin^2\mu} = \cos\mu.$$

Мы не будем напоминать все полученные в § 7 формулы, хотя многие из них нам сейчас понадобятся для получения соотношений между дугами эллипса.

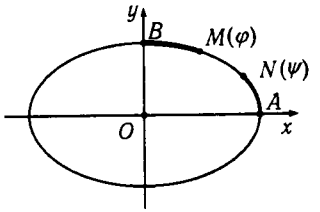


Рис. 26

Начнем с исследования случая $\mu = \pi/2$. Тогда $\sin\mu = 1$, а углы φ и ψ связаны соотношением $b\operatorname{tg}\varphi\operatorname{tg}\psi = 1$. Пусть точки M и N соответствуют углам φ и ψ (рис. 26). Обозначив через $\operatorname{arc} BM$ длину дуги BM, запишем

$$E(\varphi) = \operatorname{arc} BM,$$
$$E(\psi) = \operatorname{arc} BM + \operatorname{arc} MN,$$
$$E(\mu) = \operatorname{arc} BM + \operatorname{arc} MN + \operatorname{arc} NA.$$

Следовательно,

$$E(\varphi) + E(\psi) - E(\mu) = \operatorname{arc} BM - \operatorname{arc} NA.$$

Нами доказано следующее утверждение.

Теорема 1 (Фаньяно). *Пусть углы φ и ψ связаны соотношением $b\operatorname{tg}\varphi\operatorname{tg}\psi = 1$, M и N — соответствующие этим углам точки эллипса. Тогда разность длин дуг BM и NA равна $k^2\sin\varphi\sin\psi$.* □

Рассмотрим подробнее случай, когда точки M и N совпадают, т. е. $\varphi = \psi = \theta$. При этом $\operatorname{tg}^2\theta = b^{-1}$ и $\sin^2\theta = (1+b)^{-1}$. Следовательно,

$$BM - AM = k^2\sin^2\theta = 1 - b.$$

Кроме того, $BM + AM = E\left(\frac{\pi}{2}\right) = E^1$ — четверть длины эллипса. Поэтому

$$BM = \frac{1}{2}E^1 + \frac{1}{2}(1-b)$$

и

$$AM = \frac{1}{2}E^1 - \frac{1}{2}(1-b)$$

т. е. с точностью до алгебраической добавки $\frac{1}{2}(1-b)$ точка M делит дугу пополам. Наличие дополнительного члена

$$k^2\sin\varphi\sin\psi\sin\mu$$

в теореме сложения для эллиптических интегралов второго рода приводит к не совсем естественному делению «пополам».

Аналогичное деление дуги «пополам» можно произвести не только для четверти дуги эллипса, но и для произвольной дуги BM. Если $\varphi = \psi = \theta$ и $F(\varphi) + F(\psi) = F(\mu)$, то

$$2E(\theta) - E(\mu) = k^2\sin^2\theta\sin\mu,$$

причем углы θ и μ связаны соотношением

$$\cos^2\theta - \sin^2\theta\,\Delta(\mu) = \cos\mu.$$

Для нахождения угла μ по данному углу θ («удвоение» дуги) можно воспользоваться любой из формул

$$\sin\mu = \frac{2\sin\theta\cos\theta\,\Delta(\theta)}{1 - k^2\sin^4\theta}$$

или

$$\operatorname{tg}\left(\frac{\mu}{2}\right) = \Delta(\theta)\operatorname{tg}\theta.$$

А для нахождения угла θ по данному углу μ (деление дуги «пополам») можно воспользоваться формулой

$$\sin^2\theta = \frac{1-\cos\mu}{1+\Delta(\mu)}.$$

Равенство

$$2E(\theta) - E(\mu) = k^2\sin^2\theta\sin\mu$$

означает, что

$$2\operatorname{arc} BM - \operatorname{arc} BN = k^2\sin^2\theta\sin\mu,$$

где точки M и N соответствуют углам θ и μ. При этом

$$\operatorname{arc} BM = \frac{1}{2}\operatorname{arc} BN + \frac{1}{2}k^2\sin^2\theta\sin\mu = \frac{1}{2}\operatorname{arc} BN + \frac{1-\Delta(\mu)}{2}\operatorname{tg}\frac{\mu}{2}.$$

Таким образом, полученные формулы не позволяют построить точку, которая разбивает дугу AB ровно пополам. Однако,

построение какой-нибудь дуги MN, длина которой равна половине длины дуги AB, может быть проделано точно, без алгебраической добавки (рис. 27). Построим сначала точку K, делящую дугу AB «пополам», т. е. с алгебраической добавкой. Этой точке соответствует угол θ, для которого

$$\sin^2\theta = (1+b)^{-1}$$

и

$$E(\theta) = \frac{1}{2}E^1 + \frac{1}{2}(1-b).$$

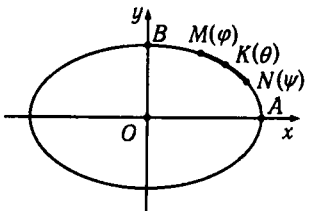


Рис. 27

Для углов φ и ψ должно выполняться соотношение

$$E(\psi) - E(\varphi) = \frac{1}{2}E^1.$$

Пользуясь свободой выбора точки M, наложим на углы φ и ψ дополнительное условие $F(\varphi) + F(\theta) - F(\psi) = 0$, из которого следует, что

$$E(\varphi) + E(\theta) - E(\psi) = k^2 \sin\varphi \sin\psi \sin\theta$$

и

$$\cos\varphi\cos\psi + \sin\varphi\sin\psi\,\Delta(\theta) = \cos\theta.$$

Тогда

$$\frac{1}{2}E^1 = E(\psi) - E(\varphi) = E(\theta) - k^2\sin\varphi\sin\psi\sin\theta =$$
$$= \frac{1}{2}E^1 + \frac{1}{2}(1-b) - k^2\sin\varphi\sin\psi\sin\theta,$$

а значит, $k^2\sin\varphi\sin\psi\sin\theta = \frac{1}{2}(1-b)$. Учитывая, что $k^2 = 1-b^2$ и $\sin^2\theta = (1+b)^{-1}$, получим

$$\sin\varphi\sin\psi = \frac{(1-b)\sqrt{1+b}}{2(1-b^2)} = \frac{1}{2\sqrt{1+b}} = \frac{1}{2}\sin\theta.$$

А так как $\Delta(\theta) = \sqrt{b}$, то соотношение

$$\cos\varphi\cos\psi + \sin\varphi\sin\psi\,\Delta(\theta) = \cos\theta$$

можно переписать в виде

$$\cos\varphi\cos\psi + \frac{\sqrt{b}}{2}\sin\theta = \cos\theta.$$

Но $\sqrt{b}\sin\theta = \cos\theta$, поэтому $\cos\varphi\cos\psi = \frac{1}{2}\cos\theta$. В результате для φ и ψ мы получаем уравнения

$$\sin\varphi\sin\psi = \frac{1}{2}\sin\theta$$

и

$$\cos\varphi\cos\psi = \frac{1}{2}\cos\theta.$$

Следовательно,

$$\cos(\varphi \pm \psi) = \frac{1}{2}(\cos\theta \mp \sin\theta) = \cos\frac{\pi}{4}\cos\left(\theta \pm \frac{\pi}{4}\right).$$

Эти формулы позволяют найти φ и ψ. Кроме того, из тех же уравнений можно получить и выражения

$$\sin\varphi = \frac{1}{4}\sqrt{3+4\sin\theta+2\sin^2\theta} - \frac{1}{4}\sqrt{3-4\sin\theta+2\sin^2\theta},$$
$$\sin\psi = \frac{1}{4}\sqrt{3+4\sin\theta+2\sin^2\theta} + \frac{1}{4}\sqrt{3-4\sin\theta+2\sin^2\theta}.$$

Найденные выражения показывают, что если дан эллипс, в котором проведены большая и малая полуоси OA и OB, то с помощью циркуля и линейки можно построить дугу эллипса, длина которой равна половине длины дуги AB. Впрочем, если полуоси не заданы, то их можно построить с помощью циркуля и линейки, хотя это и не совсем просто (см., например, [Б5, с. 114]).

Займемся теперь делением дуги эллипса на три равные части. Рассмотрим прежде всего деление с алгебраической добавкой. Напомним, что если $F(\psi) = 3F(\varphi)$, то

$$\sin\psi = \frac{3x - 4(1+k^2)x^3 + 6k^2x^5 - k^4x^9}{1 - 6k^2x^4 + 4k^2(1+k^2)x^6 - 3k^4x^8},$$

где $x = \sin\varphi$. При $\psi = \pi/2$ это уравнение сводится к уравнению

$$1 - 2x + 2k^2x^3 - k^2x^4 = 0. \qquad (2.1)$$

Если $F(\varphi_2) = 2F(\varphi)$ и $F(\varphi_3) = 3F(\varphi)$ (т. е. $\varphi_3 = \psi$), то

$$3E(\varphi) - E(\psi) = k^2\sin\varphi(\sin\varphi\sin\varphi_2 + \sin\varphi_2\sin\varphi_3).$$

Построим теперь какую-нибудь дугу, длина которой равна трети длины дуги AB, где OA и OB — полуоси эллипса. В этом случае $\psi = \varphi_3 = \pi/2$, поэтому $\sin\varphi_3 = 1$, а значит,

$$3E(\varphi) - E(\psi) = k^2\sin\varphi\sin\varphi_2(\sin\varphi + 1),$$

т. е.

$$E(\varphi) = \frac{1}{3}E^1 + \frac{1}{3}k^2 \sin\varphi \sin\varphi_2 (\sin\varphi + 1). \qquad (2.2)$$

Пусть $x = \sin\varphi$ — корень уравнения (2.1). Будем искать углы ψ и ω, удовлетворяющие соотношениям

$$E(\omega) - E(\psi) = \frac{1}{3}E^1, \qquad (2.3)$$

$$F(\varphi) + F(\psi) - F(\omega) = 0. \qquad (2.4)$$

Из (2.4) следует, что

$$E(\varphi) + E(\psi) - E(\omega) = k^2 \sin\varphi \sin\psi \sin\omega.$$

Учитывая (2.2) и (2.3), получим

$$\sin\psi \sin\omega = \frac{1}{3}\sin\varphi_2 (1 + \sin\varphi). \qquad (2.5)$$

Кроме того, из (2.4) следует, что

$$\cos\psi \cos\omega + \sin\psi \sin\omega\, \Delta(\varphi) = \cos\varphi,$$

а значит

$$\cos\psi \cos\omega = \cos\varphi - \frac{1}{3}\sin\varphi_2\, \Delta(\varphi)(1 + \sin\varphi).$$

Воспользовавшись соотношением

$$\cos\varphi_2 \cos\varphi_3 + \sin\varphi_2 \sin\varphi_3\, \Delta(\varphi) = \cos\varphi$$

получим

$$\sin\varphi_2 = \cos\varphi/\Delta(\varphi). \qquad (2.6)$$

Записав соотношение (2.1) в виде

$$1 - (1 - x^2)/(1 - k^2x^2) = (1 - x)^2,$$

где $x = \sin\varphi$, мы можем следующим образом преобразовать (2.6):

$$\cos\varphi_2 = 1 - \sin\varphi.$$

Поэтому

$$\cos\psi \cos\omega = \cos\varphi - \frac{1}{3}\cos\varphi(1 + \sin\varphi) = \frac{1}{3}\cos\varphi(1 + \cos\varphi_2).$$

Это соотношение вместе с (2.5) приводит к следующим формулам:

$$\cos(\psi \pm \omega) = \frac{1}{3}(\cos\varphi + \cos\varphi\cos\varphi_2 \mp \sin\varphi_2 \mp \sin\varphi\sin\varphi_2) =$$

$$= \frac{1}{3}(\cos\varphi \mp \sin\varphi_2 + \cos(\varphi \pm \varphi_2)).$$

Таким образом, если корень уравнения (2.1) можно построить с помощью циркуля и линейки, то для эллипса с отношением осей $1 : \sqrt{1 - k^2}$ с помощью циркуля и линейки можно построить дугу, длина которой равна трети дуги AB. Примером такого эллипса служит эллипс с отношением осей $1 : \sqrt{2}$, для которого $k^2 = 1/2$. Уравнение (2.1) в этом случае имеет вид

$$x^4 - 2x^3 + 4x - 2 = 0.$$

Сделав замену $x = \sqrt{1 - z}$, получим уравнение

$$2(1 + z)\sqrt{1 - z} = 2 - (1 - z)^2.$$

После возведения обеих частей в квадрат получим

$$z^4 + 6z^2 - 3 = 0, \quad \text{т. е.} \quad z^2 = -3 \pm 2\sqrt{3}.$$

В итоге $z = \sqrt{2\sqrt{3} - 3}$ и $\sin\varphi = x = \sqrt{1 - z}$, т. е. угол φ можно построить с помощью циркуля и линейки.

Весь материал этого параграфа взят из классического сочинения Лежандра «Traité des fonctions elliptiques» [A7].

Задачи.

1. Пусть углы φ и ψ связаны соотношением $b\,\mathrm{tg}\,\varphi\,\mathrm{tg}\,\psi = 1$. Согласно теореме Фаньяно разность дуг BM и NA равна $k^2 \sin\varphi \sin\psi$. Докажите, что она равна также отрезкам MM_1 и NN_1, где M_1 и N_1 — проекции центра эллипса на касательные в точках M и N (рис. 28).

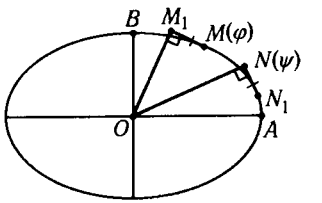


Рис. 28

§ 3. Кривые с эллиптическими дугами

Эллипс служит примером кривой, длина дуги которой выражаются эллиптическим интегралом второго рода. Для кривой, дуги которой выражаются эллиптическим интегралом первого рода, сложение дуг тоже можно было бы выполнять, причем даже без алгебраической добавки. Но есть ли вообще кривые, координаты которых являются достаточно простыми функциями параметра φ, а длина дуги, как функция аргумента φ, представляет собой эллиптический интеграл первого рода? Для $k = 1/\sqrt{2}$ такая кривая была известна еще Фаньяно, это — лемниската. Лежандр пытался построить пример такой кривой для произвольного k. Иными словами, он искал кривую $x = x(\varphi)$, $y = y(\varphi)$, длина дуги которой от точки с параметром 0 до точки с параметром φ равна

$$F(\varphi) = \int_0^{\varphi} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2\varphi}}.$$

Построить пример такой кривой не очень сложно. Если

$$dx = \frac{\cos\varphi\, d\varphi}{1 - k^2\sin^2\varphi}, \qquad dy = \frac{-b\sin\varphi\, d\varphi}{1 - k^2\sin^2\varphi},$$

где $b = \sqrt{1 - k^2}$, то

$$dx^2 + dy^2 = \frac{d\varphi^2}{1 - k^2\sin^2\varphi} = \left(\frac{d\varphi}{\Delta(\varphi)}\right)^2.$$

Поэтому в качестве функций $x(\varphi)$ и $y(\varphi)$ можно взять первообразные функций

$$\frac{\cos\varphi}{1 - k^2\sin^2\varphi}, \qquad \frac{-b\sin\varphi}{1 - k^2\sin^2\varphi}.$$

Соответствующие интегралы легко вычисляются с помощью замены переменных $u = \sin\varphi$ и $v = \cos\varphi$. В результате получаем

$$x = \frac{1}{2k}\ln\frac{1+ku}{1-ku} = \frac{1}{2k}\ln\frac{1+k\sin\varphi}{1-k\sin\varphi},$$

$$y = \frac{1}{k}\operatorname{arctg}\frac{kv}{b} = \frac{1}{k}\operatorname{arctg}\frac{k\cos\varphi}{b}.$$

Легко проверить, что

$$\cos ky = \frac{b}{\sqrt{1 - k^2\sin^2\varphi}}, \quad \operatorname{ch} kx = \frac{e^{kx} + e^{-kx}}{2} = \frac{1}{\sqrt{1 - k^2\sin^2\varphi}},$$

т. е. $\cos ky = b \operatorname{ch} kx$.

Этот пример Лежандр нашел легко, но он его не устраивал, потому что кривая $\cos ky = b\operatorname{ch} kx$ не алгебраическая. Ему удалось также построить пример алгебраической кривой, длина дуги которой равна $F(\varphi)$ с некоторой алгебраической добавкой.

Итогом этих исследований Лежандра явилась проблема нахождения всех алгебраических кривых, длина дуги которых представляем собой эллиптический интеграл первого рода. Проблему Лежандра решил французский математик Жозеф-Альфред Серре. В трех работах, опубликованных в журнале Лиувилля [А10] (изложение этих работ см. в [Б25, § 563–565; Б24]), он создал общий метод нахождения таких кривых и получил их полное описание. В частности, он построил семейство плоских алгебраических кривых S_p, зависящих от рационального положительного параметра p, длина дуги которых в точности равна $F(\varphi)$.

Кривые Серре S_p получаются следующим образом. Пусть p — фиксированное число. Рассмотрим треугольник OPM, стороны OP и PM которого равны $\sqrt{p}$ и $\sqrt{p+1}$, а углы при вершинах O и M равны α и β. Пусть

$$\cos\omega = \cos(p\alpha - (p+1)\beta). \tag{3.1}$$

Введем прямоугольную систему координат Oxy с началом в вершине O треугольника OPM так, чтобы ось Ox образовала со стороной OM угол ω (рис. 29). Будем изменять треугольник OPM так, чтобы точка O оставалась неподвижной, длины сторон OP и PM оставались постоянными, а угол ω между осью Ox и стороной OM каждый раз определялся бы соотношением (3.1). Точка M опишет при этом некоторую кривую S_p. Пусть $x = \rho\cos\omega$ и $y = \rho\sin\omega$ — координаты точки M. Можно считать, что x и y зависят от параметра α.

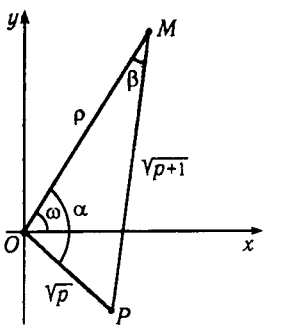


Рис. 29

Теорема 1 (Серре). *Длина дуги кривой S_p, как функция параметра α, равна*

$$\sqrt{p}\int_0^{\alpha}\frac{d\varphi}{\sqrt{1-k^2\sin^2\varphi}},\quad \text{где}\quad k=\sqrt{\frac{p}{p+1}}.$$

При этом кривая S_p алгебраическая для любого положительного рационального p.

Доказательство. Пусть $OM=\rho$. Согласно теореме косинусов

$$p+1=p+\rho^2-2\rho\sqrt{p}\cos\alpha,$$

$$p=p+1+\rho^2-2\rho\sqrt{p+1}\cos\beta,$$

т. е.

$$\cos\alpha=\frac{\rho^2-1}{2\rho\sqrt{p}},\quad \cos\beta=\frac{\rho^2+1}{2\rho\sqrt{p+1}}. \tag{3.2}$$

Таким образом, ρ и $\cos\alpha$ связаны полиномиальным соотношением. Кроме того, если $p\in\mathbb{Q}$, то $\cos\alpha$ и $\cos p\alpha$ связаны полиномиальным соотношением. Следовательно, ρ и $\cos p\alpha$ связаны полиномиальным соотношением. Аналогично ρ и $\cos(p+1)\beta$ связаны полиномиальным соотношением. А так как

$$x=\rho\cos\omega=\rho\cos(p\alpha-(p+1)\beta),$$
$$y=\rho\sin\omega,$$

то x и ρ, а также y и ρ связаны полиномиальными соотношениями. Следовательно, x и y связаны полиномиальным соотношением, т. е. кривая S_p алгебраическая.

Из (3.2) следует, что

$$\sin\alpha=\frac{R}{2\rho\sqrt{p}},\qquad \sin\beta=\frac{R}{2\rho\sqrt{p+1}},$$

где $R=\sqrt{-\rho^4+2(2p+1)\rho^2-1}$. Дифференцируя равенство

$$\cos\omega=\cos(p\alpha-(p+1)\beta),$$

получим

$$-\sin\omega\,d\omega=-(p\,d\alpha-(p+1)\,d\beta)\sin(p\alpha-(p+1)\beta).$$

При этом $\sin\omega=\pm\sin(p\alpha-(p+1)\beta)$. Следовательно,

$$\pm d\omega=p\,d\alpha-(p+1)\,d\beta.$$

Воспользовавшись тем, что

$$\cos\alpha=\frac{\rho^2-1}{2\rho\sqrt{p}}$$

и

$$d\cos\alpha=-\sin\alpha\,d\alpha=-\frac{R}{2\rho\sqrt{p}}\,d\alpha,$$

получим

$$d\alpha=-\frac{\rho^2+1}{R}\frac{d\rho}{\rho}.$$

Аналогично

$$d\beta=-\frac{\rho^2-1}{R}\frac{d\rho}{\rho}.$$

Следовательно,

$$\pm d\omega=p\,d\alpha-(p+1)\,d\beta=\frac{\rho^2-(2p+1)}{R}\frac{d\rho}{\rho}.$$

Пусть dl — дифференциал длины дуги. Тогда

$$dl^2=d\rho^2+\rho^2d\omega^2=4p(p+1)\frac{d\rho^2}{R^2},$$

т. е.

$$dl=\pm2\sqrt{p(p+1)}\frac{d\rho}{R}.$$

А так как

$$\frac{d\alpha}{\cos\beta}=-2\sqrt{p+1}\frac{d\rho}{R},$$

то

$$\pm dl=\sqrt{p}\frac{d\alpha}{\cos\beta}.$$

Кроме того, так как $\sin\beta=\sqrt{\frac{p}{p+1}}\sin\alpha$, то поэтому $\cos\beta=\sqrt{1-k^2\sin^2\alpha}$, где $k=\sqrt{\frac{p}{p+1}}$. В итоге, если отбросить знак, получим

$$dl=\sqrt{p}\frac{d\alpha}{\sqrt{1-k^2\sin^2\alpha}},$$

что и требовалось. □

Кривая S_p обладает следующим замечательным свойством. Пусть точки O, A и B соответствуют значениям параметра 0, α и β. Рассмотрим точку C, для которой длина дуги OC равна сумме длин дуг OA и OB. Иными словами, точке C соответствует

такое значение параметра γ, что $F(\gamma) = F(\alpha) + F(\beta)$. Тогда $\cos\gamma$ алгебраически выражается через $\cos\alpha$ и $\cos\beta$. Кроме того, координаты точек A, B и C алгебраически выражаются через $\cos\alpha$, $\cos\beta$ и $\cos\gamma$ соответственно. Поэтому координаты точки C алгебраически выражаются через координаты точек A и B. Таким образом, сложение и деление дуг кривой S_p — алгебраические задачи. В частности, можно написать уравнение деление дуги кривой S_p на n равных частей.

Рассмотрим теперь более подробно кривую S_p для $p = 1$. В этом случае

$$\cos\alpha = \frac{\rho^2 - 1}{2\rho}, \qquad \cos\beta = \frac{\rho^2 + 1}{2\sqrt{2}\,\rho}.$$

Несложные вычисления показывают, что

$$x = \rho\cos(\alpha - 2\beta) = \frac{\rho^4 - 1}{4\rho^2} + 1$$

и

$$y = \frac{R(1 - \rho^2)}{4\rho^2},$$

где $R = \sqrt{-\rho^4 + 6\rho^2 - 1}$. Поэтому $x - 1 = \rho_1\cos\beta$ и $y = -\rho_1\sin\beta$, где $\rho_1 = \dfrac{\rho^2 - 1}{\sqrt{2}\rho}$. Кроме того,

$$\cos 2\beta = \left(\frac{\rho^2 - 1}{2\rho}\right)^2 = \frac{1}{2}\rho_1^2,$$

т. е. рассматриваемая кривая — лемниската.

§ 4. Кривые, дуги которых выражаются через дуги окружности

В предыдущем параграфе была построена серия кривых, для которых сложение дуг — алгебраическая операция. Есть и другие примеры таких кривых. Они были построены еще Эйлером. Он писал, что нашел эти кривые лишь после длительных трудов. Координаты точек этих кривых являются алгебраическими функциями аргумента $\operatorname{tg} s$, где s — параметр, пропорциональный длине дуги кривой. Сложение дуг таких кривых есть алгебраическая операция, так как $\operatorname{tg}(\alpha + \beta)$ алгебраически выражается

через $\operatorname{tg}\alpha$ и $\operatorname{tg}\beta$. Впоследствии Серре обобщил примеры, построенные Эйлером. Он даже дал полную классификацию таких кривых. Мы ограничимся лишь наиболее простым примером.

Пусть

$$x + iy = \frac{(t - a)^{n+2}}{(t - \bar{a})^n (t + i)^2},$$

где n — целое или рациональное число, a — комплексное число. При изменении t от $-\infty$ до $+\infty$ точки с координатами (x, y) образуют некоторую кривую. Эта кривая алгебраическая, так как x и y алгебраически выражаются через t.

Легко проверить, что $dx + i\,dy$ имеет вид

$$\frac{(t - a)^{n+1}(pt + q)}{(t - \bar{a})^{n+1}(t + i)^3}\,dt.$$

Число a можно выбрать так, что $pt + q = k(t - i)$. При этом можно даже считать, что $|a| = 1$, а именно:

$$a = \frac{\sqrt{n(n+2)}}{n+1} - \frac{i}{n+1}.$$

При таком значении a получим $pt + q = k(t - i)$, где

$$k = 2\,\frac{\sqrt{n(n+2)}}{n+1}.$$

В этом случае

$$dx - i\,dy = k\,\frac{(t - \bar{a})^{n+1}(t + i)}{(t - a)^{n+1}(t - i)^3}\,dt.$$

Следовательно,

$$dl^2 = dx^2 + dy^2 = k^2\frac{dt^2}{(t + i)^2(t - i)^2} = \left(k\,\frac{dt}{t^2 + 1}\right)^2.$$

Поэтому

$$l = \pm k\int_0^t \frac{d\tau}{\tau^2 + 1} = \pm k\operatorname{arctg} t.$$

Таким образом, если параметр s равен длине дуги кривой, деленной на k, то $t = \pm\operatorname{tg} s$, причем координаты x и y точек кривой

алгебраически выражаются через t. Отсчет длин дуг ведется от точки, соответствующей параметру $t = 0$.

Задачи

1. Докажите, что при $n = 1$ после гомотетии уравнение рассматриваемой кривой в полярных координатах можно записать в виде

$$\cos\varphi = \frac{\rho^2 + 6\rho - 2}{3\rho^2\sqrt{3}}.$$

ГЛАВА 4

ТЕОРЕМА АБЕЛЯ О ДЕЛЕНИИ ЛЕМНИСКАТЫ

Лемнискатой называют кривую, уравнение которой в полярных координатах имеет вид $r^2 = \cos 2\theta$ (рис. 30). Это название происходит от латинского слова lemniscatus — украшенный лентами. В декартовых координатах (x, y), где $x = r\cos\theta$, $y = r\sin\theta$, эта кривая имеет уравнение

$$(x^2 + y^2)^2 = x^2 - y^2.$$

В самом деле, $x^2 + y^2 = r^2$ и $x^2 - y^2 = r^2\cos 2\theta$.

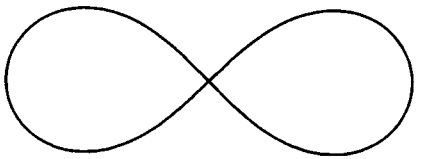


Рис. 30

Первым к исследованию лемнискаты обратился французский астроном Жак Доминик (Джованни Доминико) Кассини (1625–1712), итальянец по национальности. Он рассматривал даже более общие кривые, для точек которых произведение расстояний до двух фиксированных точек F_1 и F_2 постоянно (рис. 31). На основании астрономических наблюдений Кассини полагал, что с помощью таких кривых движение планет можно описать более точно, чем с помощью эллипсов. Эти кривые называют теперь овалами Кассини. Но книга Кассини «Eléments d'astronomie» («Основы астрономии»), в которой они изучались, была опубликована лишь через много лет после его смерти,

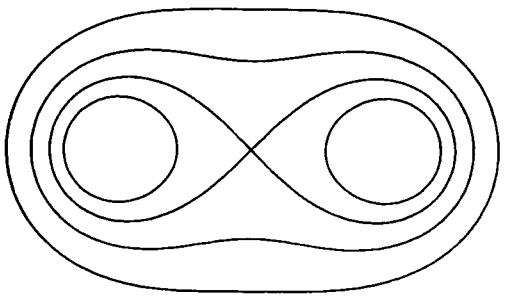


Рис. 31

в 1749 г. Математическому сообществу лемниската стала известна из статей Якоба Бернулли и Иоганна Бернулли, опубликованных в 1694 г. [A2]. Поэтому ее обычно называют лемнискатой Бернулли.

Наиболее замечательные свойства лемнискаты были обнаружены итальянским математиком графом Фаньяно (1682–1766). Кстати сказать, именно он ввел термин «эллиптические интегралы». Фаньяно обнаружил, что длина дуги лемнискаты выражается эллиптическим интегралом первого рода. Он получил теорему сложения для этого интеграла и тем самым показал, что деление дуги лемнискаты на n равных частей — это алгебраическая задача. В 1750 г. Фаньяно издал собрание своих статей под названием «Produzioni matematiche» [A11]. Берлинская академия поручила написать отзыв об этой книге Леонарду Эйлеру. Работы Фаньяно вызвали у него большой интерес к эллиптическим интегралам. В своих многочисленных исследованиях Эйлер существенно развил и обобщил методы и результаты Фаньяно.

Уже Фаньяно было известно, что деление лемнискаты сводится к решению алгебраического уравнения. Но методы анализа разрешимости уравнений в квадратных радикалах тогда еще не были разработаны. Первым существенного успеха в этой области добился в 1796 г. 19-летний Гаусс. Он обнаружил, что правильный 17-угольник можно построить с помощью циркуля и линейки, т. е. уравнение $x^{17} - 1 = 0$ разрешимо в квадратных радикалах. Позже Гаусс доказал, что с помощью циркуля и линейки можно построить правильный n-угольник для всех n вида $2^a p_1 \dots p_k$, где p_i — различные простые числа Ферма, т. е. простые числа вида $2^{2^m} + 1$. Гаусс писал, что для всех остальных n правильный n-угольник нельзя построить с помощью циркуля и линейки, но нет никаких подтверждений того, что он действительно умел это доказывать.

Уравнение деления лемнискаты тоже интересовало Гаусса. Например, он показал, что уравнение 25-й степени, связанное с делением лемнискаты на 5 равных частей, решается в квадратных радикалах. Его рассуждения были основаны на том, что число 5 можно представить в виде произведения двух комплексно сопряженных чисел $2 + i$ и $2 - i$ (см. § 3). Этих исследований Гаусс не опубликовал, но в книге «Disquisitiones Arithmeticae» («Арифметические исследования»), изданной в 1801 г. [A3, a], он упомянул, что разработанные им методы применимы не только

к тригонометрическим функциям, но и к функциям, связанным с интегралами вида $\int \frac{dx}{\sqrt{1-x^4}}$. Это утверждение заинтриговало Абеля. Он подробно исследовал уравнение деления лемнискаты и доказал, что лемнискату можно разделить на n равных частей для всех чисел n вида $2^a p_1 \dots p_k$, где p_i — различные простые числа Ферма. Эту теорему Абель считал одним из наиболее важных своих достижений. Она содержится во второй части его большой работы «Recherches sur les fonctions elliptiques» («Исследования по эллиптическим функциям») [A1]. Доказательство Абеля весьма длинное и сложное. Впоследствии Эйзенштейн (1823–1852) получил более простое доказательство. При этом он обнаружил интересные свойства многочленов деления лемнискаты.

Недавно Роузен [B14] нашел новое изящное доказательство теоремы Абеля. Оно сравнительно несложно, и, кроме того, ясно показывает решающее значение инвариантности решетки периодов лемнискатических функций относительно умножения на i. Чтобы читатель мог почувствовать стиль эпохи Абеля и, в то же время, познакомиться с современным истолкованием предмета, мы приведем оба доказательства — и то, которое полтора века назад получил Эйзенштейн, и то недавнее, которое было найдено Роузеном.

Абель доказал лишь возможность деления лемнискаты на n равных частей с помощью циркуля и линейки при указанных значениях n. Однако, он не доказывал, что при других n этого сделать нельзя. И, тем не менее, как показано в работе [B14], при других n нельзя построить с помощью циркуля и линейки координаты точек, делящих лемнискату на n равных частей. Но это не означает, что при других n нельзя разделить лемнискату на на n равных частей с помощью циркуля и линейки в том случае, когда лемниската нарисована. Дело в том, что использование самой лемнискаты дает для построений дополнительные возможности. Рассматривая точки пересечения прямых и окружностей с лемнискатой можно, вообще говоря, строить не только квадратичные иррациональности.

И последнее замечание. Известно, что с помощью циркуля и линейки можно разделить пополам любую дугу окружности. Оказывается, что с помощью циркуля и линейки можно разделить пополам и любую дугу лемнискаты. Доказательство этого факта приведено на с. 113.

Прежде чем заняться уравнением деления лемнискаты, рассмотрим более простое уравнение деления окружности. Сначала мы покажем, как вполне элементарным способом можно решить в квадратных радикалах уравнение $x^{17}-1=0$, хотя это решение не переносится на уравнение деления лемнискаты. Затем обсудим подход к разрешимости уравнения $x^n-1=0$ в квадратных радикалах, который переносится и на уравнение деления лемнискаты.

§ 1. Построение правильного 17-угольника. Элементарный подход

Корни уравнения $x^n-1=0$ являются вершинами правильного n-угольника. В самом деле, если $\varepsilon=e^{2\pi i/n}$, то $\varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1}$ — корни рассматриваемого уравнения.

Поделив многочлен $x^n-1=0$ на $x-1$, получим многочлен $x^{n-1}+x^{n-2}+\ldots+x+1$. Таким образом, если уравнение

$$x^{n-1}+x^{n-2}+\ldots+x+1=0 \qquad (1.1)$$

разрешимо в квадратных радикалах, то правильный n-угольник можно построить с помощью циркуля и линейки.

При $n=3$ никаких вопросов не возникает, потому что квадратное уравнение $x^2+x+1=0$, безусловно, разрешимо в квадратных радикалах. При $n=5$ уравнение (1.1) решается легко. В самом деле, после замены $u=x+x^{-1}$ его можно переписать в виде $u^2+u-1=0$.

При $n=17$ решить уравнение (1.1) в квадратных радикалах уже не так просто. Гаусс использовал для этого следующий метод, применимый и к более сложным случаям. Попытаемся расположить корни 17-й степени из единицы ε, ε^2, ε^3, ..., ε^{16}, где $\varepsilon=e^{2\pi i/17}$ в виде цикла так, чтобы каждый из них получался из предыдущего одним и тем же способом, а именно, возведением в одну и ту же степень. Если возводить в квадрат, то таким образом можно получить только восемь корней:

$$\varepsilon,\ \varepsilon^2,\ \varepsilon^4,\ \varepsilon^8,\ \varepsilon^{16},\ \varepsilon^{15},\ \varepsilon^{13},\ \varepsilon^9,$$

так как $(\varepsilon^9)^2$ опять равно ε.

Попробуем возводить ε в некоторую более высокую степень g. Тогда получается ряд чисел

$$\varepsilon,\quad \varepsilon^g,\quad \varepsilon^{g^2},\quad \varepsilon^{g^3},\quad \ldots,\quad \varepsilon^{g^{15}}.$$

Мы должны подобрать такое g, чтобы выполнялось следующее условие: остатки от деления чисел $1, g, g^2, \ldots, g^{15}$ на 17 должны принимать значения от 1 до 16. Такие числа называются *примитивными корнями по модулю* 17. Более общим образом, целое число a называется *примитивным корнем по простому модулю* p, если $s=p-1$ — наименьшее целое положительное число, для которого $a^s\equiv 1 \pmod p$.

Гаусс доказал существование примитивных корней для любого заданного простого числа. К сожалению, не существует простого способа нахождения таких корней. Для небольших простых чисел метод подбора, по-видимому, ничем не хуже любого другого метода. В частности, для $p=17$ имеется восемь примитивных корней: $g=3,5,6,7,10,11,12,14$.

Возьмем примитивный корень $g=3$. В результате получим последовательность чисел $\varepsilon_k=\varepsilon^{3^k}$, $k=0,\ldots,15$:

$$\varepsilon,\ \varepsilon^3,\ \varepsilon^9,\ \varepsilon^{10},\ \varepsilon^{13},\ \varepsilon^5,\ \varepsilon^{15},\ \varepsilon^{11},\ \varepsilon^{16},\ \varepsilon^{14},\ \varepsilon^8,\ \varepsilon^7,\ \varepsilon^4,\ \varepsilon^{12},\ \varepsilon^2,\ \varepsilon^6.$$

Пусть x_1 — сумма чисел ε_k с четными номерами k, x_2 — сумма чисел ε_k с нечетными номерами k, т. е.

$$x_1=\varepsilon+\varepsilon^9+\varepsilon^{13}+\varepsilon^{15}+\varepsilon^{16}+\varepsilon^8+\varepsilon^4+\varepsilon^2,$$
$$x_2=\varepsilon^3+\varepsilon^{10}+\varepsilon^5+\varepsilon^{11}+\varepsilon^{14}+\varepsilon^7+\varepsilon^{12}+\varepsilon^6.$$

Сумма всех корней уравнения $x^{17}-1=0$, включая корень $x=1$, равна нулю, поэтому $x_1+x_2=-1$. Несложные вычисления показывают, что $x_1x_2=-4$. В самом деле, пусть $\alpha=2\pi/17$. Тогда $\varepsilon^k=\cos k\alpha+i\sin k\alpha$, поэтому $\varepsilon+\varepsilon^{16}=2\cos\alpha$, $\varepsilon^9+\varepsilon^8=2\cos 8\alpha$, $\varepsilon^{13}+\varepsilon^4=2\cos 4\alpha$ и $\varepsilon^{15}+\varepsilon^2=2\cos 2\alpha$, т. е.

$$x_1=2(\cos\alpha+\cos 8\alpha+\cos 4\alpha+\cos 2\alpha).$$

Аналогично

$$x_2=2(\cos 3\alpha+\cos 7\alpha+\cos 5\alpha+\cos 6\alpha).$$

Воспользовавшись формулой

$$2\cos p\alpha\cos q\alpha=\cos(p+q)\alpha+\cos(p-q)\alpha,$$

получим

$$x_1x_2 = 8(\cos\alpha+\cos 2\alpha+\cos 3\alpha+\ldots+\cos 8\alpha) = 4(x_1+x_2) = -4.$$

Таким образом, x_1 и x_2 можно найти, решив квадратное уравнение

$$x^2 + x - 4 = 0 \qquad (1.2)$$

Так как

$$\cos\alpha + \cos 2\alpha > 2\cos\pi/4 = \sqrt{2} > -\cos 8\alpha$$

и $\cos 4\alpha > 0$, то $x_1 > 0$. Поэтому $x_2 = -4/x_1 < 0$, т. е. x_1 — положительный корень уравнения (1.2), а x_2 — отрицательный.

Обозначим суммы чисел ε_k с номерами, дающими при делении на 4 остатки 0, 1, 2 и 3, через y_1, y_3, y_2 и y_4 соответственно. Тогда

$$y_1 = \varepsilon+\varepsilon^{13}+\varepsilon^{16}+\varepsilon^4 = 2(\cos\alpha+\cos 4\alpha),$$
$$y_2 = \varepsilon^9+\varepsilon^{15}+\varepsilon^8+\varepsilon^2 = 2(\cos 8\alpha+\cos 2\alpha),$$
$$y_3 = \varepsilon^3+\varepsilon^5+\varepsilon^{14}+\varepsilon^{12} = 2(\cos 3\alpha+\cos 5\alpha),$$
$$y_4 = \varepsilon^{10}+\varepsilon^{11}+\varepsilon^7+\varepsilon^6 = 2(\cos 7\alpha+\cos 6\alpha).$$

Ясно, что $y_1+y_2 = x_1$ и $y_1 > y_2$, так как $\cos\alpha > \cos 2\alpha$ и $\cos 4\alpha > \cos 8\alpha$. Кроме того,

$$y_1y_2 = 2(\cos\alpha+\ldots+\cos 8\alpha) = -1.$$

Таким образом, y_1 и y_2 удовлетворяют уравнению $y^2 - x_1y - 1 = 0$.

Легко проверить, что y_3 и y_4 удовлетворяют уравнению $y^2 - x_2y - 1 = 0$, причем $y_3 > y_4$.

Рассмотрим, наконец, $z_1 = \varepsilon+\varepsilon^{16} = 2\cos\alpha$ и $z_2 = \varepsilon^{13}+\varepsilon^4 = 2\cos 4\alpha$, т. е. суммы чисел ε_k с номерами, дающими при делении на 8 остатки 0 и 4. Тогда $z_1 > z_2$, $z_1+z_2 = y_1$ и

$$z_1z_2 = 4\cos\alpha\cos 4\alpha = 2(\cos 5\alpha+\cos 3\alpha) = y_3.$$

Поэтому z_1 — больший корень уравнения $z^2 - y_1z + y_3 = 0$. Таким образом, отрезок длиной $z_1 = 2\cos(2\pi/17)$ можно построить с помощью циркуля и линейки. После этого правильный 17-угольник строится очевидным образом.

Для полноты, следуя [Б32], приведем явное описание построения 17-угольника.

1) Восстановим перепендикуляр к прямой UV в точке O и отложим на нем отрезок OA, равный единице, а из точки O по UV — отрезок $OC = 1/4$. Из C, как из центра, проведем окружность радиуса CA, пересекающую UV в точках B и D.

Тогда $OB = x_1/2$, $OD = -x_2/2$.

2) Соединим A с B, и из B, как из центра, проведем окружность радиуса BO, пересекающую AB и ее продолжение в точках P и M. Тогда $AM = y_1$, $AP = -y_2$. Аналогично получим, что $AN = y_3$ и $AQ = -y_4$.

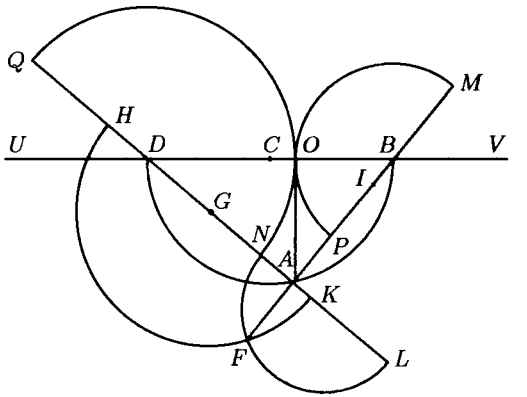


Рис. 32

3) На продолжении AD отложим $AL = NO$. На NL как на диаметре построим окружность, которая пересекается с продолжением AB в точке F. Из F, как из центра, проведем окружность радиуса $AI = AM/2$. Эта окружность пересекает AD в точке G. Из G, как из центра, проведем окружность того же радиуса, которая пересечет AD в точках K и H. Тогда $AH = z_1$ и $AK = z_2$. Таким образом, отрезок AK равен стороне правильного 34-угольника, вписанного в круг радиуса единица.

§ 2. Построение правильных многоугольников. Элементы теории Галуа

В предыдущем параграфе было показано, как можно решить в квадратных радикалах уравнение $x^{17} - 1 = 0$. Теперь мы докажем, что для всех чисел n вида $2^np_1\ldots p_k$, где p_i — различные простые числа Ферма, уравнение $x^n - 1 = 0$ разрешимо в квадратных радикалах. При этом наше изложение будет таким, чтобы его почти без изменений можно было перенести на случай лемнискаты.

Сопоставив каждому действительному числу t точку с координатами $(\cos t, \sin t)$, получим параметризацию единичной ок-

ружности C действительными числами. В результате C превращается в абелеву группу с единичным элементом $(1,0)$. Так как

$$\cos(t+s) = \cos t \cos s - \sin t \sin s,$$
$$\sin(t+s) = \sin t \cos s + \cos t \sin s,$$

то закон сложения точек окружности записывается следующим образом:

$$(a,b)+(c,d) = (ac-bd, ad+bc) = (f(a,b,c,d), g(a,b,c,d)).$$

Легко проверить, что

$$2(x,y) = (x^2-y^2, 2xy)$$

и

$$3(x,y) = (x^3-3xy^2, 3x^2y-y^3).$$

Аналогично $n(x,y) = (f_n(x,y), g_n(x,y))$, где f_n и g_n — многочлены с целыми коэффициентами. Из соотношения $\cos n\varphi + i\sin n\varphi = (\cos\varphi + i\sin\varphi)^n$ получаем

$$f_n(x,y) = \frac{(x+iy)^n + (x-iy)^n}{2},$$
$$g_n(x,y) = \frac{(x+iy)^n - (x-iy)^n}{2i}. \qquad (2.1)$$

Пусть C_n — множество таких точек $(x,y) \in C$, что $n(x,y) = (1,0)$, т. е. $f_n(x,y) = 1$ и $g_n(x,y) = 0$. Эти точки служат вершинами правильного n-угольника. Ясно также, что C_n — подгруппа C, причем она изоморфна $\mathbb{Z}/n\mathbb{Z}$ — аддитивной группе вычетов по модулю n.

Над полем $\mathbb{C}$ решениями системы уравнений $f_n(x,y) = 1$, $g_n(x,y) = 0$ будут не только точки C_n. Найдем все эти решения. Воспользовавшись формулами (2.1), можно перейти к эквивалентной системе уравнений $(x+iy)^n = 1$, $(x-iy)^n = 1$. Следовательно, $x+iy = \varepsilon^p$ и $x-iy = \varepsilon^q$, где $\varepsilon = e^{2\pi i/n}$. В частности, $x^2+y^2 = \varepsilon^p\varepsilon^q$, поэтому равенство $x^2+y^2 = 1$ выполняется тогда и только тогда, когда $\varepsilon^p = \varepsilon^{-q}$. Таким образом, C_n можно охарактеризовать как множество всех решений системы уравнений

$$\begin{cases} f_n(x,y) = 1, \\ g_n(x,y) = 0, \\ x^2+y^2 = 1. \end{cases} \qquad (2.2)$$

Рассмотрим поле K_n, порожденное над $\mathbb{Q}$ координатами всех точек множества C_n. Например, C_3 состоит из точек $(1,0)$ и $(-1/2, \pm\sqrt{3}/2)$, поэтому $K_3 = \mathbb{Q}(\sqrt{3})$. Множество C_4 состоит из точек $(\pm 1, 0)$ и $(0, \pm 1)$, поэтому $K_4 = \mathbb{Q}$. Пусть σ — автоморфизм поля K_n, оставляющий элементы поля $\mathbb{Q}$ неподвижными. Так как коэффициенты многочленов f_n и g_n — целые числа, то σ определяет некоторую перестановку точек C_n. По этой перестановке автоморфизм σ однозначно восстанавливается, так как координаты точек C_n порождают поле K_n. Ясно также, что автоморфизму σ соответствует не произвольная перестановка элементов группы C_n, а ее автоморфизм. Следовательно, группа G_n автоморфизмов поля K_n над $\mathbb{Q}$ изоморфна некоторой подгруппе группы $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$.

Дальнейший ход доказательства таков. Сначала мы покажем, что если $n = 2^a p_1 \dots p_k$, где p_i — различные простые числа Ферма, то порядок группы $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ равен степени двойки. Тогда, в частности, порядок группы G_n тоже равен степени двойки. Затем мы докажем, что если порядок группы G равен 2^k, то существует такая последовательность подгрупп

$$G = G^0 \supset G^1 \supset \dots \supset G^k = \{e\},$$

что G^i — подгруппа G^{i-1} индекса 2 для $i = 1, \dots, k$. Наконец, по этой последовательности подгрупп мы построим последовательность квадратичных расширений полей, начинающуюся полем $\mathbb{Q}$ и заканчивающуюся полем K_n. Существование такой последовательности расширений означает, что все элементы поля K_n, в частности, и координаты точек C_n являются квадратичными иррациональностями.

Лемма 1. *Порядок группы* $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ *равен степени двойки тогда и только тогда, когда* $n = 2^a p_1 \dots p_k$, *где* p_i — *различные простые числа Ферма.*

Доказательство. Автоморфизмы аддитивной группы $\mathbb{Z}/n\mathbb{Z}$ имеют вид $x \longmapsto mx$, где m — число, взаимно простое с n. Поэтому порядок группы $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ равен $\varphi(n)$, где $\varphi(n)$ — количество чисел, не превосходящих n и взаимно простых с n.

Если числа p и q взаимно просты, то $\varphi(pq) = \varphi(p)\varphi(q)$. В самом деле, пусть $0 \leqslant a \leqslant p-1$ и $0 \leqslant b \leqslant q-1$. Тогда остатки от деления $n = pq$ чисел вида $aq+bp$ на n образуют полную систему вычетов по модулю n. Для доказательства этого достаточно

заметить, что

$$a_1 q + b_1 p \equiv a_2 q + b_2 p \pmod{pq}$$

тогда и только тогда, когда

$$(a_1 - a_2) q \equiv (b_2 - b_1) p \pmod{pq},$$

т. е. $a_1 \equiv a_2 \pmod{p}$ и $b_1 \equiv b_2 \pmod{q}$. Ясно также, что числа $aq + bp$ и pq взаимно просты тогда и только тогда, когда числа a и p, а также b и q взаимно просты.

Если $n = p^k$, где p — простое число, то $\varphi(n) = p^{k-1}(p-1)$. В самом деле, среди чисел, не превосходящих n, общий делитель с n имеют лишь числа $p, 2p, \ldots, p^{k-1}p$. Их количество равно p^{k-1}.

Пусть $n = p_1^{k_1} \ldots p_m^{k_m}$. Тогда $\varphi(n)$ есть произведение чисел $p_i^{k_i-1}(p_i-1)$. Число $p^{k-1}(p-1)$ может быть степенью двойки лишь в двух случаях:

а) $p = 2$, при этом k — любое;

б) $p - 1 = 2^c$, при этом $k = 1$.

Во втором случае число c не может иметь нечетных делителей. В самом деле, если d — нечетный делитель c, то $2^c + 1 = x^d + 1$ делится на $x + 1$. Следовательно, p — простое число вида $2^{2^l} + 1$. □

Лемма 2. *Если порядок группы G равен 2^k, то существует такая последовательность подгрупп*

$$G = G^0 \supset G^1 \supset \ldots \supset G^k = \{e\},$$

что G^i — подгруппа G^{i-1} индекса 2 для $i = 1, \ldots, k$.

Доказательство. Применим индукцию по k. При $k = 1$ утверждение очевидно. Предположим, что утверждение доказано для всех групп порядка 2^{k-1}.

Прежде всего докажем, что в группе G существует не единичный элемент, коммутирующий со всеми остальными. Для каждого элемента $x \in G$ рассмотрим класс сопряженных с ним элементов, т. е. множество элементов вида gxg^{-1}, где $g \in G$. Любые два класса либо совпадают, либо не пересекаются, т. е. группа G разбита на попарно не пересекающиеся классы сопряженных элементов. Равенство $g_1 x g_1^{-1} = g_2 x g_2^{-1}$ эквивалентно равенству $xh = hx$, где $h = g_2^{-1} g_1$. Рассмотрим подгруппу

$$G_x = \{h \in G \mid xh = hx\}.$$

Элементы $g_1 x g_1^{-1}$ и $g_2 x g_2^{-1}$ равны тогда и только тогда, когда $g_1 \in g_2 G_x$. Поэтому число элементов класса $\{gxg^{-1}\}$ равно индексу подгруппы G_x в G, а значит, оно имеет вид 2^s.

Класс $\{gxg^{-1}\}$ содержит ровно один элемент лишь в том случае, когда x коммутирует со всеми элементами группы G, т. е. x — элемент центра G. Предположим, что центр группы G состоит лишь из единичного элемента. Тогда сумма мощностей всех классов сопряженных элементов равна $1 + 2^{s_1} + \ldots + 2^{s_p}$, где $s_i \geqslant 1$. Следовательно, эта сумма нечетна. С другой стороны, она равна порядку группы G, т. е. равна 2^k. Получено противоречие, поэтому центр группы G содержит неединичный элемент a. Он порождает циклическую подгруппу порядка 2^r. Рассмотрим элемент $b = a^m$, где $m = 2^{r-1}$. Подгруппа H, порожденная элементом b, имеет порядок 2 и является нормальной, поскольку b лежит в центре группы G. По предположению индукции для группы $F = G/H$ порядка 2^{k-1} существует последовательность

$$F = F^0 \supset F^1 \supset \ldots \supset F^{k-1} = \{e\},$$

где F^i — подгруппа F^{i-1} индекса 2 для $i = 1, \ldots, k-1$. Чтобы получить требуемую последовательность подгрупп G, положим $G^k = H$ и $G^i = F^i \cup bF^i$ при $i = 0, \ldots, k-1$. □

Пусть $n = 2^a p_1 \ldots p_m$, где p_i — различные простые числа Ферма. Тогда у группы G_n автоморфизмов поля K_n над $\mathbb{Q}$ есть последовательность подгрупп

$$G = G^0 \supset G^1 \supset \ldots \supset G^k = \{e\},$$

где G^i — подгруппа G^{i-1} индекса 2 для $i = 1, \ldots, k$. Сопоставим подгруппе G^i множество L_i, состоящее из тех элементов поля K_n, которые остаются неподвижными при действии всех автоморфизмов из G_i. Сумма, разность, произведение и отношение элементов L_i принадлежит L_i, поэтому L_i — поле. Ясно, что $L_k = K_n$ и $L_i \supset L_{i-1}$.

Все автоморфизмы из G^i оставляют элементы поля L_i неподвижными. Кроме того, $G^{i-1} = G^i \cup \sigma G^i$, где σ — любой элемент из $G^{i-1} \setminus G^i$. Поэтому $\sigma^2 \in G^i$, так как множеству σG^i этот элемент принадлежать не может. Следовательно, автоморфизм σ поля K_n таков, что если $x \in L_i$, то $\sigma^2 x = x$. Кроме того, $\sigma x = x$ тогда и только тогда, когда $x \in L_{i-1}$. Любой элемент $x \in L_i$

можно представить в виде суммы элементов $x_1 = (x + \sigma x)/2$ и $x_2 = (x - \sigma x)/2$. При этом $\sigma x_1 = x_1$ и $\sigma x_2 = -x_2$. Следовательно, $x_1 \in L_{i-1} \subset L_i$ и $x_2 = x - x_1 \in L_i$. Предположим, что $L_i \neq L_{i-1}$. Пусть $a \in L_i \setminus L_{i-1}$ и $\alpha = \sigma a - a$. Тогда $\sigma\alpha = -\alpha$, причем $\alpha \neq 0$. Кроме того, $\sigma(\alpha x_2) = (-\alpha)(-x_2) = \alpha x_2$, т. е. $\alpha x_2 \in L_{i-1}$ и $x_2 \in \alpha^{-1} L_{i-1}$. Таким образом, $L_i = L_{i-1} + \alpha^{-1} L_{i-1}$. Другими словами, размерность L_i над L_{i-1} равна двум. Следовательно, если $x \in L_i$, то элементы 1, x, x^2 линейно зависимы над полем L_{i-1}. Поэтому $x^2 + px + q = 0$, где $p, q \in L_{i-1}$. Значит, любой элемент поля K_n — квадратичная иррациональность над полем L_0.

Докажем теперь, что $L_0 = \mathbb{Q}$. Для этого достаточно доказать, что для любого $x \in K_n \setminus \mathbb{Q}$ найдется автоморфизм поля K_n над $\mathbb{Q}$, сдвигающий x. Вообще говоря, не любое расширение поля $\mathbb{Q}$ обладает таким свойством. Например, у поля

$$\{p + q\sqrt[3]{2} + r\sqrt[3]{4} \mid p, q, r \in \mathbb{Q}\}$$

нет автоморфизмов, отличных от тождественного. В самом деле, элемент $\sqrt[3]{2}$ может перейти лишь в корень уравнения $x^3 - 2 = 0$, а рассматриваемому полю принадлежит только один корень этого уравнения.

Заметим сначала, что K_n — это алгебраическое расширение $\mathbb{Q}$, т. е. координаты точек C_n — алгебраические числа. В самом деле, решения системы уравнений $f_n(x, y) = 1$, $g_n(x, y) = 0$ алгебраические. Для доказательства достаточно рассмотреть $f_n(x, y) - 1$ и $g_n(x, y)$ как многочлены от y и записать их результант.

Пусть $\overline{\mathbb{Q}}$ — множество всех алгебраических над $\mathbb{Q}$ чисел. Легко проверить, что $\overline{\mathbb{Q}}$ — поле. В самом деле, пусть α, $\beta \in \overline{\mathbb{Q}}$. Тогда

$$\alpha^p = a_0 + a_1\alpha + \ldots + a_{p-1}\alpha^{p-1},$$

$$\beta^q = b_0 + b_1\beta + \ldots + b_{q-1}\beta^{q-1},$$

где $a_i, b_i \in \mathbb{Q}$, причем $a_0 b_0 \neq 0$. Поэтому любой элемент кольца, порожденного над $\mathbb{Q}$ элементами α и β, можно представить в виде линейной комбинации с рациональными коэффициентами элементов $\alpha^i\beta^j$, где $0 \leqslant i < p$ и $0 \leqslant j < q$. В частности, элементы 1, $\alpha + \beta$, ..., $(\alpha + \beta)^{pq}$ можно представить в виде линейной комбинации указанных pq элементов, поэтому они линейно зависимы над полем $\mathbb{Q}$, т. е. $\alpha + \beta \in \overline{\mathbb{Q}}$. Аналогично доказывается, что

$\alpha\beta \in \overline{\mathbb{Q}}$. Кроме того, $a_0\alpha^{-1} = \alpha^{p-1} - a_1 - \ldots - a_{p-1}\alpha^{p-2}$, поэтому $\alpha^{-1} \in \overline{\mathbb{Q}}$.

Любой автоморфизм τ поля $\overline{\mathbb{Q}}$ над $\mathbb{Q}$ переводит поле K_n в себя. В самом деле, поле K_n порождено координатами точек C_n, а C_n совпадает с множеством всех решений над $\mathbb{C}$ системы уравнений (2.2).

Равенство $L_0 = \mathbb{Q}$ означает, что если $a \in K_n \setminus \mathbb{Q}$, то существует автоморфизм σ: $K_n \to K_n$, для которого $\sigma(a) \neq a$. Для доказательства этого достаточно указать автоморфизм τ: $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}$, для которого $\tau(a) \neq a$.

Теорема 1. *Пусть a и b — корни одного и того же неприводимого многочлена над $\mathbb{Q}$. Тогда существует автоморфизм τ: $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}$, для которого $\tau(a) = b$.*

Доказательство. Пусть K — поле, α — корень неприводимого многочлена P на полем K, $K(\alpha)$ — поле, порожденное α над K. Тогда произвольный изоморфизм f: $K \to K'$ можно продолжить до изоморфизма g: $K(\alpha) \to K'(\beta)$, где β — корень многочлена $f(P)$. В самом деле, поле $K(\alpha)$ состоит из элементов вида $\sum k_j\alpha^j$, где $j \geqslant 0$ и $k_j \in K$. Положим $g\left(\sum k_j\alpha^j\right) = \sum f(k_j)\beta^j$. Это определение корректно, так как равенство $\sum k_j\alpha^j = 0$ эквивалентно тому, что многочлен $F = \sum k_j\alpha^j$ делится на P.

Построим сначала изоморфизм τ_1: $\mathbb{Q}(a) \to \mathbb{Q}(b)$. Выберем затем элемент $t_2 \in \overline{\mathbb{Q}} \setminus \mathbb{Q}(a)$. Он является корнем неприводимого многочлена P_2 над $\mathbb{Q}(a)$. Пусть t_2' — корень многочлена $\tau_1(P_2)$. Тогда можно построить изоморфизм τ_2: $\mathbb{Q}(a, t_2) \to \mathbb{Q}(b, t_2')$. Выберем элемент $t_3 \in \overline{\mathbb{Q}} \setminus (a, t_2)$ и построим изоморфизм τ_3: $\mathbb{Q}(a, t_2, t_3) \to \mathbb{Q}(b, t_2', t_3')$ и т. д. Так как размерность пространства $\overline{\mathbb{Q}}$ над $\mathbb{Q}$ счетна, то можно построить базис $\{1, \varepsilon_1 = a, \varepsilon_2, \varepsilon_3, \ldots\}$ пространства $\overline{\mathbb{Q}}$ над $\mathbb{Q}$. Элементы t_2, t_3, ... можно выбирать так, чтобы поле $\mathbb{Q}(a, t_2, \ldots, t_k)$ содержало подпространство, порожденное элементами 1, ε_1, ..., ε_k. В результате мы построим мономорфизм τ: $\overline{\mathbb{Q}} \to \overline{\mathbb{Q}}$, для которого $\tau(a) = b$. Остается проверить, что τ — эпиморфизм. Пусть $\gamma_1 \in \overline{\mathbb{Q}}$ — корень неприводимого многочлена R над $\mathbb{Q}$, $\gamma_1, \ldots, \gamma_n$ — все корни этого многочлена. Тогда $\tau(\gamma_i) \in \{\gamma_1, \ldots, \gamma_n\}$, причем числа $\tau(\gamma_1), \ldots, \tau(\gamma_n)$ попарно различны. Поэтому множества $\{\gamma_1, \ldots, \gamma_n\}$ и

$\{\tau(\gamma_1),\dots,\tau(\gamma_n)\}$ совпадают. В частности, $\gamma_1 = \tau(\gamma_j)$ для некоторого j. □

Теорему 1 можно существенно обобщить. Заметим сначала, что поле $\overline{\mathbb{Q}}$ алгебраически замкнуто. В самом деле, пусть x_0 — корень многочлена $\alpha + \beta x + \dots + \omega x^n = 0$, где $\alpha, \beta, \dots, \omega \in \overline{\mathbb{Q}}$. Рассмотрим неприводимый над $\mathbb{Q}$ многочлен, имеющий корень α. Пусть $\alpha_1,\dots,\alpha_p$ — корни этого многочлена. Тогда все элементарные симметрические функции от $\alpha_1,\dots,\alpha_p$ выражаются через его коэффициенты, поэтому они рациональны. Определим аналогично $\beta_1,\dots,\beta_q$; …; $\omega_1,\dots,\omega_r$ и рассмотрим многочлен

$$P(x) = \prod_{i,j,\dots,k} (\alpha_i + \beta_j x + \dots + \omega_k x^n).$$

Этот многочлен ненулевой, причем все его коэффициенты выражаются через $\sigma_s(\alpha_1,\dots,\alpha_p),\dots,\sigma_t(\omega_1,\dots,\omega_r)$, поэтому они рациональны. А так как $P(x_0) = 0$, то $x_0 \in \overline{\mathbb{Q}}$. □

В общем случае для автоморфизмов алгебраически замкнутого поля Ω над подполем K справедливо следующее утверждение.

Теорема 2. *Пусть x, $y \in \Omega$, где Ω — какое-либо расширение поля K. Тогда автоморфизм поля Ω над K, переводящий x в y, существует тогда и только тогда, когда либо x и y оба трансцендентные над K, либо оба они алгебраические и являются корнями одного и того же неприводимого многочлена над K.*

Для полей общего вида эту теорему мы доказывать не будем (доказательство см. [Б9]).

Но для наиболее интересного случая — автоморфизмов $\mathbb{C}$ над $\mathbb{Q}$ — мы приведем доказательство не только этой теоремы, но и некоторых ее обобщений. Например, мы докажем, что мощность множества автоморфизмов $\mathbb{C}$ совпадает с мощностью множества всех отображений $\mathbb{C} \to \mathbb{C}$ (т. е. она больше мощности $\mathbb{C}$). Наше изложение следует статьям [В7, В16].

Для начала заметим, что изоморфизм полей $\varphi\colon F \to G$ продолжается до изоморфизма $\varphi'\colon F(\alpha) \to G(\beta)$ тогда и только тогда, когда выполняются следующие условия:

1) если элемент α алгебраичен над F и P — неприводимый многочлен над F с корнем α, то β — корень многочлена $\varphi(P)$;

2) если элемент α трансцендентен над F, то β трансцендентен над G.

Для дальнейших рассуждений нам понадобится лемма Цорна. Дело в том, что доказательства по индукции применимы лишь к счетным множествам, а размерность $\mathbb{C}$ над $\mathbb{Q}$ несчетна. Поэтому для работы с автоморфизмами $\mathbb{C}$ над $\mathbb{Q}$ требуется другой аппарат, и лемма Цорна достаточно удобна для этих целей.

Прежде чем сформулировать лемму Цорна, дадим некоторые определения. Пусть M — некоторое множество. Обозначим через 2^M множество всех его подмножеств. Множество $A \subset 2^M$ называют *цепью*, если для любых двух его элементов $a, b \in A$ либо $a \subset b$, либо $b \subset a$. Множество $B \subset 2^M$ называют *замкнутым по Цорну*, если для любой цепи $A \subset B$ множество B содержит также объединение всех элементов цепи A. Элемент $m \in B$ называют *максимальным*, если множество $m \subset M$ не содержится ни в каком другом множестве $a \subset M$, являющемся элементом B.

Лемма (Цорн). *Каждое непустое замкнутое по Цорну множество $B \subset 2^M$ содержит по крайней мере один максимальный элемент m.*

С помощью леммы Цорна можно построить продолжение автоморфизма φ подполя в $\mathbb{C}$ до автоморфизма всего поля $\mathbb{C}$. Для этого нужно применить лемму Цорна к семейству автоморфизмов, продолжающих φ. Но тут есть одна небольшая трудность. Пусть $\mathbb{Q} \subset F \subset F'$ — поля и φ — автоморфизм F над $\mathbb{Q}$. Тогда продолжение φ до автоморфизма F' над $\mathbb{Q}$ существует отнюдь не всегда. Например, пусть $F = \mathbb{Q}(\sqrt{2})$, $F' = \mathbb{Q}(\sqrt[4]{2})$ и φ задан формулой: $a + b\sqrt{2} \mapsto a - b\sqrt{2}$. Тогда $\sqrt[4]{2} \in F'$ под действием автоморфизма должен переходить в элемент $\tau \in F'$, квадрат которого равен $\tau^2 = -\sqrt{2}$. Легко видеть, что элементов τ с таким свойством в поле F' нет.

Однако, если поле F' является алгебраическим замыканием поля F, то следующее утверждение показывает, что продолжение автоморфизма φ существует.

Теорема 3. *Изоморфизм полей $\varphi\colon F \to G$ можно продолжить до изоморфизма $\overline{F} \to \overline{G}$ их алгебраических замыканий.*

Доказательство. Рассмотрим всевозможные продолжения автоморфизма φ до изоморфизма $\varphi_\alpha\colon F_\alpha \to G_\alpha$, где

$F_\alpha \subset \overline{F}$, а значит, $G_\alpha \subset \overline{G}$. Покажем, что множество всех подмножеств в $\overline{F} \times \overline{G}$ вида

$$\{(a, \varphi_\alpha(a)) \mid a \in F_\alpha\}$$

замкнуто по Цорну. Рассмотрим в этом множестве произвольную цепь. По определению цепи для любых двух ее элементов соответствующие им изоморфизмы φ_α и φ_β таковы, что один из этих изоморфизмов является продолжением другого. Это означает, что объединению всех элементов цепи соответствует некоторый изоморфизм, т. е. их объединение принадлежит рассматриваемому множеству.

Согласно лемме Цорна рассматриваемое множество имеет максимальный элемент. Этому элементу соответствует изоморфизм $\psi\colon F' \to G'$. Нужно доказать, что $F' = \overline{F}$ и $G' = \overline{G}$. Предположим, что некоторый элемент $a \in \overline{F}$ не принадлежит F'. Элемент a алгебраичен над полем F', а поле $\overline{G}$ алгебраически замкнуто. Поэтому поле $\overline{G}$ содержит элемент b, являющийся корнем образа при изоморфизме ψ минимального многочлена элемента a над полем F'. Следовательно, изоморфизм $\psi\colon F' \to G'$ можно продолжить до изоморфизма $F'(a) \to G'(b)$, что противоречит максимальности элемента, соответствующего изоморфизму ψ.

Итак, $F' = \overline{F}$. Остается доказать, что $G' = \overline{G}$. Поле G' изоморфно $\overline{F}$, поэтому оно алгебраически замкнуто. Кроме того, поле G' содержит G. Следовательно, $G' = \overline{G}$. □

Теперь можно доказать теорему о продолжении автоморфизмов подполей поля $\mathbb{C}$.

Теорема 4. *Любой автоморфизм ψ подполя в $\mathbb{C}$ продолжается до автоморфизма всего поля $\mathbb{C}$.*

Доказательство. Рассмотрим всевозможные продолжения данного автоморфизма $\varphi\colon F \to F$ до автоморфизмов $\varphi_\alpha\colon F_\alpha \to F_\alpha$, где $F_\alpha \subset \mathbb{C}$. Как и при доказательстве теоремы 3, получаем, что множество, состоящее из множеств вида

$$\{(a, \varphi_\alpha(a)) \mid a \in F_\alpha\},$$

имеет максимальный элемент. Этому элементу соответствует автоморфизм $\varphi'\colon F' \to F'$. Нужно доказать, что $F' = \mathbb{C}$. Пусть некоторое комплексное число a не принадлежит F'. Если число a алгебраично над F', то согласно теореме 3 существует продолжение автоморфизма φ' до автоморфизма алгебраического замыкания поля F', которое строго больше F'. Если же элемент a трансцендентен над полем F', то существует продолжение изоморфизма φ' до изоморфизма $F'(a) \to F'(a)$, переводящего a в a. В обоих случаях получаем противоречие с максимальностью элемента F'. Следовательно, $F' = \mathbb{C}$. □

Отметим, что изоморфизм двух подполей в $\mathbb{C}$ не всегда можно продолжить до автоморфизма поля $\mathbb{C}$. Например, существует изоморфизм $\mathbb{C} \to F \subset \mathbb{C}$, где $F \neq \mathbb{C}$. Такой изоморфизм строится следующим образом. Пусть $a_1, a_2, \ldots$ — счетное множество комплексных чисел, алгебраически независимых над $\mathbb{Q}$. Отображение $a_i \mapsto a_{i+1}$ задает изоморфизм

$$\mathbb{Q}(a_1, a_2, \ldots) \to \mathbb{Q}(a_2, a_2, \ldots) \subset \mathbb{Q}(a_1, a_2, \ldots).$$

Рассмотрим всевозможные продолжения этого изоморфизма до таких изоморфизмов $\Theta_\alpha\colon F_\alpha \to G_\alpha$ ($F_\alpha, G_\alpha \subset \mathbb{C}$), что элемент a_1 трансцендентен над G_α. По лемме Цорна множество $\{F_\alpha\}$ имеет максимальный элемент, который, как легко показать, совпадает с $\mathbb{C}$. Таким образом, получаем изоморфизм $\mathbb{C} \to F \subset \mathbb{C}$, где поле F не содержит числа a_1, а потому $F \neq \mathbb{C}$.

Доказательство теоремы 2 в случае автоморфизмов поля $\mathbb{C}$ над полем $\mathbb{Q}$ теперь не представляет трудностей. В самом деле, если комплексные числа x и y трансцендентны, то можно рассмотреть автоморфизм поля $\mathbb{Q}(x, y)$, меняющий местами x и y. Согласно теореме 4 этот автоморфизм можно продолжить до автоморфизма поля $\mathbb{C}$. Если же числа x и y являются корнями одного и того же неприводимого многочлена над полем $\mathbb{Q}$, то существует изоморфизм $\mathbb{Q}(x) \to \mathbb{Q}(y)$, приводящий x в y. Согласно теореме 3 его можно продолжить до изоморфизма алгебраических замыканий полей $\mathbb{Q}(x)$ и $\mathbb{Q}(y)$. Но алгебраические замыкания этих полей совпадают, поэтому мы теперь получаем автоморфизм поля $\overline{\mathbb{Q}(x)} = \overline{\mathbb{Q}(y)}$, который по теореме 4 продолжается до автоморфизма поля $\mathbb{C}$.

Теперь уже ясно, что мощность автоморфизмов поля $\mathbb{C}$ не меньше мощности континуума. Но оказывается, что мощность этого множества больше мощности континуума.

Теорема 5. *Мощность множества всех автоморфизмов поля $\mathbb{C}$ совпадает с мощностью всех отображений $\mathbb{C}$ в $\mathbb{C}$.*

Доказательство. Нужно доказать, что мощность множества всех автоморфизмов поля $\mathbb{C}$ совпадает с мощностью множества всех подмножеств континуального множества. При этом

достаточно доказать, что мощность множества всех автоморфизмов поля $\mathbb{C}$ не меньше мощности множества всех подмножеств континуального множества.

Множество $B \subset \mathbb{C}$ называют *базисом трансцендентности* над $\mathbb{Q}$, если B алгебраически независимо над $\mathbb{Q}$ и B не содержится в другом множестве комплексных чисел, алгебраически независимых над Q. Из максимальности множества B следует, что поле $\mathbb{C}$ алгебраично над полем $\mathbb{Q}(B)$. Поэтому, в частности, множество B имеет мощность континуума.

Покажем, что любому подмножеству S в B можно сопоставить автоморфизм φ_S поля $\mathbb{C}$ так, чтобы разным множествам соответствовали разные автоморфизмы. Рассмотрим автоморфизм поля $\mathbb{Q}(B)$ над полем $\mathbb{Q}$, переводящий элемент $x \in B$ в элемент x, если $x \in S$, и в элемент $-x$, если $x \notin S$. Согласно теореме 4 этот автоморфизм можно продолжить до автоморфизма всего поля $\mathbb{C}$. Ясно, что если элемент x принадлежит одному из множеств S и T, а другому не принадлежит, то $\varphi_S(x) = -\varphi_T(x)$, поэтому $\varphi_S \neq \varphi_T$. □

Задачи

1. Докажите, что любой автоморфизм поля $\mathbb{R}$ тождествен.

Указание. Любой автоморфизм σ поля $\mathbb{R}$ должен оставлять неподвижными элементы поля $\mathbb{Q}$. Кроме того, неравенство $x \geqslant y$ эквивалентно тому, что $x - y = a^2$, $a \in \mathbb{R}$.

2. Пусть f — неприводимый многочлен над $\mathbb{Q}$, имеющий корень $\cos(2k\pi/n)$. Докажите, что все корни многочлена f вещественны.

Указание. Любой автоморфизм поля $\overline{\mathbb{Q}}$ над полем $\mathbb{Q}$ сохраняет поле $K_n \subset \mathbb{R}$.

§ 3. Уравнение деления лемнискаты

Вычислим длину дуги лемнискаты $r^2 = \cos 2\theta$. Напомним, что дифференциал дуги кривой в полярных координатах можно вычислить по формуле $ds^2 = d\tau^2 + r^2\,d\theta^2$. Действительно, если $x = r\cos\theta$ и $y = r\sin\theta$, то

$$ds^2 = dx^2 + dy^2 = (\cos\theta\,dr - r\sin\theta\,d\theta)^2 + \\ + (\sin\theta\,dr + r\cos\theta\,d\theta)^2 = dr^2 + r^2 d\theta^2.$$

В нашем случае $2r\,dr = -2\sin\theta\,d\theta$.

Поэтому

$$ds^2 = dr^2 + \frac{r^4 dr^2}{1 - \cos^2 2\theta} = \frac{dr^2}{1 - r^4}.$$

Следовательно,

$$s(r) = \int\limits_0^r \frac{dx}{\sqrt{1 - x^4}} = \int\limits_0^r \frac{dx}{\sqrt{(1 - x^2)(1 - k^2x^2)}}, \quad \text{где} \quad k^2 = -1.$$

Таким образом, $r = \operatorname{sn} s$ — эллиптический синус Якоби для $k^2 = -1$. Правда, с таким модулем k синус Якоби обычно не рассматривают, но теорема сложения, безусловно, остается справедливой:

$$\operatorname{sn}(u + v) = \frac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v + \operatorname{sn} v \operatorname{cn} u \operatorname{dn} u}{1 + \operatorname{sn}^2 u\, sn^2 v},$$

где $\operatorname{cn} u = \sqrt{1 - \operatorname{sn}^2 u}$ и $\operatorname{dn} u = \sqrt{1 + \operatorname{sn}^2 u}$.

В дальнейшем будем обозначать $\operatorname{sn} s$ для $k^2 = -1$ через $\varphi(s)$. Соотношение

$$ds = \frac{dr}{\sqrt{1 - r^4}} = \frac{d\varphi(s)}{\sqrt{1 - \varphi^4(s)}}$$

означает, что $\varphi' = \sqrt{1 - \varphi^4}$. Поэтому теорему сложения можно записать в виде

$$\varphi(u + v) = \frac{\varphi(u)\,\varphi'(v) + \varphi(v)\,\varphi'(u)}{1 + \varphi^2(u)\,\varphi^2(v)}.$$

Функция φ обладает очень важным свойством, которое нам неоднократно придется использовать: $\varphi(iu) = i\varphi(u)$. В самом деле, после замены $x = iy$ получим

$$\int\limits_0^{ir} \frac{dx}{\sqrt{1 - x^4}} = i \int\limits_0^r \frac{dy}{\sqrt{1 - y^4}}.$$

Следовательно, если

$$u = \int\limits_0^r \frac{dy}{\sqrt{1 - y^4}},$$

то $r = \varphi(u)$ и $ir = \varphi(iu)$. Из соотношения $\varphi(iu) = i\varphi(u)$ следует,

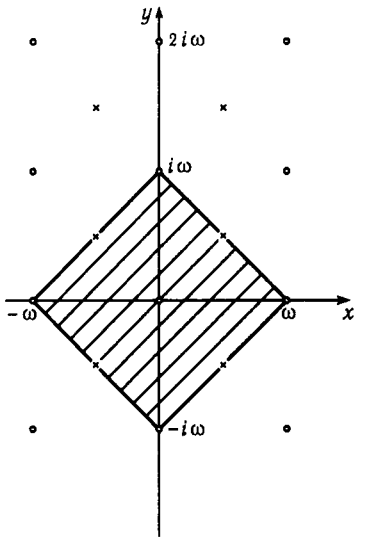


Рис. 33

что $\varphi'(iu) = \varphi'(u)$ и $\varphi'(-u) = \varphi'(u)$.

Соотношение $\varphi(iu) = i\varphi(u)$ означает, что решетка периодов функции φ переходит в себя при отображении $u \longmapsto iu$. Выясним, как именно устроена эта решетка.

Пусть

$$\frac{\omega}{2} = \int_0^1 \frac{dx}{\sqrt{1-x^4}}.$$

Тогда длина лемнискаты равна 2ω. Число ω играет для лемнискаты такую же роль, как число π для окружности. Согласно определению $\varphi(\omega/2) = 1$ и $\varphi'(\omega/2) = 0$. Из свойств функции sn следует, что $\varphi(u+\omega) = -\varphi(u)$. Поэтому $\varphi(u + i\omega) = -\varphi(u)$, а значит, $\varphi(u+\omega \pm i\omega) = \varphi(u)$, т. е. $\omega(1 \pm i)$ — периоды функции φ.

Найдем теперь нули и полюсы функции φ. Пусть $\alpha, \beta \in \mathbb{R}$. Тогда

$$\varphi(\alpha + i\beta) = \frac{\varphi(\alpha)\varphi'(\beta) + i\varphi'(\alpha)\varphi(\beta)}{1 - \varphi^2(\alpha)\varphi^2(\beta)}.$$

Заметим, что φ и φ' не имеют на действительной оси полюсов. Поэтому равенство $\varphi(\alpha + i\beta) = 0$ возможно лишь в том случае, когда $\varphi(\alpha)\varphi'(\beta) = \varphi'(\alpha)\varphi(\beta) = 0$. Вещественные нули функций φ и φ' имеют соответственно вид $m\omega$ и $(m+1/2)\omega$, где $m \in \mathbb{Z}$. Поэтому нули функций φ должны иметь вид $m\omega + ni\omega$ или $(m+1/2)\omega + (n+1/2)i\omega$. Очевидно, что $\varphi(m\omega + ni\omega) = 0$. Равенство

$$\varphi(u+\omega/2)\varphi(u+i\omega/2) = \frac{\varphi'(u)}{1+\varphi^2(u)} \frac{i\varphi'(u)}{1-\varphi^2(u)} = i$$

показывает, что если $\varphi(u+\omega/2) = 0$, то $\varphi\left(u + i\omega/2\right) = \infty$. Поэтому $(m+1/2)\omega + (n+1/2)i\omega$ — полюсы функции φ. Система нулей и полюсов функции φ изображена на рис. 33 (полюсы отмечены крестиками), фундаментальный параллелограмм заштрихован.

Обратимся теперь к задаче деления лемнискаты на равные части. Пусть длина дуги лемнискаты, заключенной между началом координат и точкой (r, θ), равна s. Тогда $r = \varphi(s)$. Длина одного лепестка лемнискаты равна ω, поэтому точки, для которых $s = k\omega/n$, где $0 < k < n$, делят его на n равных частей (рис. 34). Чтобы построить эти точки, достаточно построить отрезки длиной $r = \varphi(\alpha)$, где $\alpha = k\omega/n$. В самом деле $\cos 2\theta = r^2$, поэтому точку (r, θ) можно построить с помощью циркуля и линейки, если известен отрезок r. Так как $n\alpha = k\omega$, то $\varphi(n\alpha) = 0$. Теорема сложения для функции φ позволяет выразить $\varphi(n\alpha)$ через $\varphi(\alpha)$, т. е. $\varphi(n\alpha) = F_n(\varphi)$, где $\varphi = \varphi(\alpha)$, а F_n — алгебраическая функция.

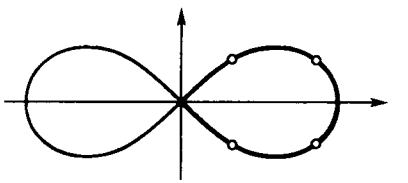


Рис. 34

Таким образом, задача деления на n равных частей дуги лемнискаты, заключенной между началом координат и точкой (r, θ), сводится к решению уравнения $F_n(\varphi) = 0$. В частности, для того, чтобы разделить на n равных частей весь лепесток лемнискаты достаточно решить уравнение $F_n(\varphi) = 0$.

Прежде, чем обратиться к общему случаю, рассмотрим уравнение $F_n(\varphi) = r$ при некоторых n.

Начнем со случая $n = 2$. Для того, чтобы разделить пополам с помощью циркуля и линейки дугу лемнискаты, нужно решить уравнение $F_2(\varphi) = r$ в квадратичных радикалах. Согласно теореме сложения

$$F_2(\varphi) = \varphi(2\alpha) = \frac{2\varphi(\alpha)\varphi'(\alpha)}{1+\varphi^4(\alpha)} = \frac{2\varphi\sqrt{1-\varphi^4}}{1+\varphi^4}.$$

Уравнение

$$2\varphi\sqrt{1-\varphi^4} = r(1+\varphi^4)$$

после возведения в квадрат и замены $x = \varphi^2 - \varphi^{-2}$ перепишется в виде

$$x^2 + 4r^{-2}x + 4 = 0.$$

Покажем, что можно разделить пополам дугу лемнискаты, заключенную между произвольными точками (r_1, θ_1) и (r_2, θ_2). Пусть искомая точка имеет координаты (r_0, θ_0). Ясно, что $r_0 = \varphi\left(\frac{s_1+s_2}{2}\right)$, где $r_i = \varphi(s_i)$, $i = 1, 2$. Построение r_0 сведется к

решению уравнения $F_2(\varphi) = r$, если мы сумеем построить отрезок $\varphi(s_1 + s_2)$. Для этого достаточно воспользоваться теоремой сложения

$$\varphi(s_1 + s_2) = \frac{\varphi(s_1)\sqrt{1-\varphi^4(s_2)} + \varphi(s_2)\sqrt{1-\varphi^4(s_1)}}{1+\varphi^2(s_1)\varphi^2(s_2)}.$$

Пусть $n = 3$. Так как

$$\varphi(2\alpha+\alpha) + \varphi(2\alpha-\alpha) = \frac{2\,\varphi(2\alpha)\,\varphi'(\alpha)}{1+\varphi^2(2\alpha)\,\varphi^2(\alpha)},$$

то

$$\varphi(3\alpha) = -\varphi\,\frac{\varphi^8+6\varphi^4-3}{1+6\varphi^4-3\varphi^8}.$$

Таким образом, деление лемнискаты на три равные части сводится к решению уравнения $\varphi^8 + 6\varphi^4 - 3 = 0$. Оно, очевидно, разрешимо в квадратных радикалах. Наконец, пусть $n = 5$. Для вычисления $\varphi(5\alpha)$ можно воспользоваться тем, что

$$\varphi(3\alpha+2\alpha) + \varphi(3\alpha-2\alpha) = \frac{2\,\varphi(3\alpha)\,\varphi'(2\alpha)}{1+\varphi^2(3\alpha)\,\varphi^2(2\alpha)}.$$

Несложные, но громоздкие вычисления показывают, что

$$\varphi(5\alpha) = \varphi\,\frac{\varphi^{24}+50\varphi^{20}-125\varphi^{16}+300\varphi^{12}-105\varphi^8-62\varphi^4+5}{1+50\varphi^4-125\varphi^8+300\varphi^{12}-105\varphi^{16}-62\varphi^{20}+5\varphi^{24}}. \quad (3.1)$$

Для решения уравнения $F_5(\varphi) = 0$ в квадратных радикалах Гаусс воспользовался тем, что над кольцом $\mathbb{Z}[i]$ число 5 разлагается на множители $2+i$ и $2-i$. Пусть $\beta = (2+i)\,\alpha$ и $\overline{\beta} = (2-i)\,\alpha$. Тогда

$$\varphi(\beta) = \varphi\,\frac{-i\varphi^4+2+i}{1-(1-2i)\,\varphi^4} = \psi, \qquad \varphi(\overline{\beta}) = \varphi\,\frac{i\varphi^4+2-i}{1-(1+2i)\,\varphi^4} = \overline{\psi},$$

$$\varphi(5\alpha) = \psi\,\frac{i\psi^4+2-i}{1-(1+2i)\,\psi^4}, \qquad \varphi(5\alpha) = \overline{\psi}\,\frac{-i\overline{\psi}^4+2+i}{1-(1-2i)\,\overline{\psi}^4}.$$

Отметим, кстати, что числитель дроби (3.1) делится на числители дробей ψ и $\overline{\psi}$. Поделив его на

$$(-i\varphi^4+2+i)\,(i\varphi^4+2-i) = \varphi^8 - 2\varphi^4 + 5,$$

получим многочлен

$$\varphi^{16} + 52\varphi^{12} - 26\varphi^8 - 12\varphi^4 + 1.$$

Ненулевое решение уравнения $\varphi(5\alpha) = 0$, можно получить, решив сначала уравнение $i\psi^4 + 2 - i = 0$, а затем уравнение

$$\varphi\,\frac{-i\varphi^4+2+i}{1-(1-2i)\,\varphi^4} = \psi = \sqrt[4]{1+2i}. \quad (3.2)$$

Можно также сначала решить уравнение $-i\overline{\psi}^4+2+i = 0$, а затем уравнение

$$\varphi\,\frac{i\varphi^4+2-i}{1-(1+2i)\,\varphi^4} = \overline{\psi} = \sqrt[4]{1-2i}. \quad (3.3)$$

Поделив (3.2) на (3.3), получим квадратное уравнение относительно φ^4.

Перейдем к общему случаю. Заметим, что формула сложения для φ и соотношение $\varphi(i\alpha) = i\varphi(\alpha)$ позволяют выразить величину $\varphi((a+bi)\alpha)$, $a, b \in \mathbb{Z}$, через $\varphi(\alpha)$:

$$\varphi((a+bi)\alpha) = F_{a+bi}(\alpha).$$

Вообще говоря, функция F_{a+bi} алгебраическая. Но если $a+b$ — нечетное число, то эта функция рациональна. Чтобы проверить это, заметим сначала, что

$$\psi = \varphi(\alpha \pm i\alpha) = (1 \pm i)\,\frac{\varphi}{\sqrt{1-\varphi^4}}$$

и

$$\varphi'(\alpha \pm i\alpha) = \sqrt{1-\psi^4} = \frac{1+\varphi^4}{1-\varphi^4},$$

при этом в последней формуле знак корня определяется тем, что $\varphi'(0) = 1$. Кроме того, сложив выражения для $\varphi(x+y)$ и $\varphi(x-y)$, получим

$$\varphi(x+y) + \varphi(x-y) = \frac{2\,\varphi(x)\,\varphi'(y)}{1+\varphi^2(x)\,\varphi^2(y)}.$$

Пусть $y = \alpha \pm i\alpha$. Тогда

$$\varphi(x+y) + \varphi(x-y) = \frac{2\,\varphi(x)\,(1+\varphi^4)}{1-\varphi^4 \pm 2i\,\varphi^2(x)\,\varphi^2}.$$

Таким образом, если $\varphi(x)$ и $\varphi(x-y)$ — рациональные функции от φ, то $\varphi(x+y)$ тоже рациональная функция от φ. Например,

положив $x = \alpha$, получим

$$\varphi(2\alpha + i\alpha) = \varphi\,\frac{-i\varphi^4 + 2 + i}{1 - (1-2i)\,\varphi^4},$$

$$\varphi(2\alpha - i\alpha) = \varphi\,\frac{i\varphi^4 + 2 - i}{1 - (1+2i)\,\varphi^4}.$$

Рассуждения по индукции показывают, что если число $a + b$ нечетно, то $\varphi((a+bi)\alpha) = \varphi\dfrac{P(\varphi^4)}{Q(\varphi^4)}$, где P и Q — многочлены с коэффициентами из кольца $\mathbb{Z}[i] = \{p + qi \,|\, p, q \in \mathbb{Z}\}$. Если при этом $b = 0$, то коэффициенты многочленов P и Q — целые числа.

Будем называть комплексное число $a + bi$ нечетным, если a и b — целые числа, и при этом $a + b$ нечетно.

Докажем теперь некоторые свойства многочленов P и Q для произвольных нечетных чисел $m = a + bi$. Любое такое число m, прибавляя к нему числа вида $\pm 2\,(1 \pm i)$, можно привести к виду ± 1 или $\pm i$, т. е. к виду i^ε.

Теорема 1. *Справедливо равенство:*

$$\varphi(m\alpha) = i^\varepsilon \varphi\,\frac{\varphi^{p-1} + A_1\varphi^{p-5} + \ldots + A_{\frac{p-1}{4}}}{1 + A_1\varphi^4 + \ldots + A_{\frac{p-1}{4}}\varphi^{p-1}},$$

где $\varphi = \varphi(\alpha)$ *и* $p = a^2 + b^2$ — *квадрат нормы числа* m.

Доказательство. Воспользуемся соотношением

$$\varphi\left(u + i^\varepsilon\omega/2\right)\varphi\left(u + i^{\varepsilon+1}\omega/2\right) = (-1)^\varepsilon i.$$

Пусть $x = \varphi\left(u + \omega/2\right)$ и $y = \varphi\left(u + i\omega/2\right)$. Тогда $xy = i$ и

$$\varphi\left(mu + i^\varepsilon\omega/2\right)\varphi\left(mu + i^{\varepsilon+1}\omega/2\right) = (-1)^\varepsilon i,$$

т. е.

$$x\,\frac{P(x)}{Q(x)}\,y\,\frac{P(y)}{Q(y)} = (-1)^\varepsilon i.$$

Следовательно,

$$\frac{P(x)}{Q(x)}\,\frac{P(ix^{-1})}{Q(ix^{-1})} = (-1)^\varepsilon.$$

Кроме того,

$$P(ix^{-1}) = P(x^{-1}) \quad \text{и} \quad Q(ix^{-1}) = Q(x^{-1}),$$

так как P и Q зависят лишь от x^4. Поэтому многочлены P и Q имеют одну и ту же степень r, причем $Q(x) = \lambda x^r P(x^{-1})$. Коэффициент λ можно найти, воспользовавшись тем, что $\varphi\left(\omega/2\right) = 1$ и $\varphi\left(m\omega/2\right) = i^\varepsilon$. В самом деле, это означает, что $i^\varepsilon = \dfrac{P(1)}{\lambda P(1)}$, т. е. $\lambda = i^{-\varepsilon}$.

Остается показать, что $r = a^2 + b^2 - 1$. При преобразовании $z \mapsto z/m$ решетка, порожденная ω и $i\omega$, переходит в решетку нулей функции $\varphi(m\alpha)$. Поэтому нули функции $\varphi(m\alpha)$ образуют решетку с площадью фундаментального параллелограмма $\omega^2\,|m|^{-2}$. Площадь параллелограмма периодов функции $\varphi(\alpha)$ равна $2\omega^2$, поэтому он содержит $2\,|m|^2$ нулей функции $\varphi(m\alpha) = \varphi\dfrac{P(\varphi)}{Q(\varphi)}$. Так как степени многочлена P и Q равны, то равенство $\varphi(m\alpha) = 0$ выполняется тогда и только тогда, когда выполняется одно из равенств $\varphi = 0$ или $P(\varphi) = 0$. Каждое уравнение $\varphi(\alpha) = a$ имеет в параллелограмме периодов ровно два решения, поэтому степень многочлена P равна $\dfrac{2|m|^2 - 2}{2} = |m|^2 - 1$. □

Корни уравнения $P(\varphi) = 0$ при нечетном комплексном m могут быть описаны следующим образом. Пусть $\Omega = \omega(1 - i)$ и $\Omega' = \omega(1 + i)$ — периоды функции $\varphi(\alpha)$. Из равенства $\varphi(m\alpha) = 0$ следует, что

$$m\alpha = (a + bi)\,\omega = p\Omega + p'\Omega',$$

где $p = a/2 - b/2$, $p' = a/2 + b/2$. так как $\Omega' = 2\Omega$, то $m\alpha = (p + ip')\Omega$. Обозначая комплексное число $p + ip' \in \mathbb{Z}[i]$ через ν, получим, что $m\alpha = \nu\Omega$, т. е. $\alpha = \nu\Omega/m$. Заставим ν пробегать полную систему вычетов по модулю m, за исключением нуля. Тогда числа $\varphi(\nu\Omega/m)$ будут корнями уравнения $P(\varphi) = 0$. Все эти корни различны между собой. В самом деле, равенство

$$\varphi\left(\frac{\nu\Omega}{m}\right) = \varphi\left(\frac{\nu'\Omega}{m}\right)$$

возможно лишь в двух случаях: или когда

$$\nu \equiv \nu' \pmod m$$

что исключено, или же когда

$$2(\nu + \nu') = (1 + i)\,(1 - 2i\nu)\,m$$

что также невозможно, поскольку правая часть делится только на $1+i$, а левая часть — на 2. Таким образом, все корни уравнения $P(\varphi)=0$ имеют вид

$$x_\nu = \varphi\left(\frac{\nu\Omega}{m}\right),$$

где ν пробегает полную систему вычетов по модулю m, исключая нулевой вычет.

На примере вычисления $\varphi(m\alpha)$ при $m=5$ мы уже видели, что важны арифметические свойства m не столько над кольцом $\mathbb{Z}$, сколько над кольцом $\mathbb{Z}[i]$. Это подтверждает и следующая теорема, доказанная Эйзенштейном.

Теорема 2. *Пусть $m=a+bi$ — нечетное простое над $\mathbb{Z}[i]$ число. Тогда числа $A_1, \ldots, A_{\frac{p-1}{4}}$ делятся на m.*

Доказательство. Поделив многочлен

$$A_{\frac{p-1}{4}} + A_{\frac{p-5}{4}}\varphi^4 + \ldots + \varphi^{p-1}$$

на многочлен

$$1 + A_1\varphi^4 + \ldots + A_{\frac{p-1}{4}}\varphi^{p-1},$$

получим

$$\varphi(m\alpha) = \varphi(c_0 + c_1\varphi^4 + c_2\varphi^8 + \ldots), \qquad (3.4)$$

где $c_j \in \mathbb{Z}[i]$, причем $c_0 = A_{\frac{p-1}{4}}$. Другое выражение для $\varphi(m\alpha)$ можно получить, воспользовавшись тем, что

$$\varphi'(\alpha) = \sqrt{1-\varphi^4(\alpha)} = 1 + \sum s_j\varphi^{4j}(\alpha), \qquad (3.5)$$

где $s_j \in \mathbb{Q}$. В самом деле, так как $\varphi(0)=0$ и $\varphi'(0)=1$, то

$$\varphi(\alpha) = \alpha\left(1 + \sum p_j\alpha^j\right).$$

Равенство $\varphi'^2 = 1-\varphi^4$ может выполняться лишь в том случае, когда $p_j=0$ при $j \not\equiv 0 \pmod 4$. Значит

$$\varphi(\alpha) = \alpha(1 + d_1\alpha^4 + d_2\alpha^8 + \ldots),$$

где $d_j \in \mathbb{Q}$. Заменив α на $m\alpha$, получим

$$\varphi(m\alpha) = m\alpha(1 + d_1(m\alpha)^4 + d_2(m\alpha)^8 + \ldots). \qquad (3.6)$$

Чтобы сравнить (3.6) с (3.4), заметим еще, что

$$\varphi^k = \varphi^k(\alpha) = \alpha^k(1 + d_1\alpha^4 + d_2\alpha^8 + \ldots)^k.$$

Поэтому (3.4) можно переписать в виде

$$\varphi(m\alpha) = c_0\alpha(1 + d_1\alpha^4 + d_2\alpha^8 + \ldots) + \\ + c_1\alpha^5(1 + e_1\alpha^4 + e_2\alpha^8 + \ldots) + \ldots \qquad (3.7)$$

Сравнивая выражения (3.6) и (3.7), получим

$$\begin{aligned} m &= c_0, \\ d_1m^5 &= c_0d_1 + c_1, \\ d_2m^9 &= c_0d_2 + c_1e_1 + c_2, \\ d_3m^{13} &= c_0d_3 + c_1e_2 + c_2f_1 + c_3 \end{aligned}$$

и т. д. Следовательно, $c_k = mH_k(m)l^{-1}$, где H_k — многочлен степени $4k$ с целыми коэффициентами, в совокупности взаимно простыми с целым числом l. Формула $nH_k(n)l^{-1} = c_k(n)$ справедлива для любого нечетного $n \in \mathbb{Z}[i]$ (не обязательно простого); при этом $c_k(n) \in \mathbb{Z}[i]$. Следовательно, $nH_k(n) \equiv 0 \pmod l$.

Пусть $q \in \mathbb{Z}[i]$ — простой нечетный делитель числа l. Тогда $nH_k(n) \equiv 0 \pmod q$ для любого нечетного числа n. С другой стороны, сравнение $xH_k(x) \equiv 0 \pmod q$ не может иметь более $\deg(xH_k(x)) = 4k+1$ различных по модулю q решений. Для нечетного числа q остатки от деления нечетных чисел на q образуют полную систему вычетов, так как если число n_1 четно, то число $n = n_1 + q$ нечетно. Решетка чисел, пропорциональных q, порождена векторами q и iq. Площадь фундаментального параллелограмма этой решетки равна $|q|^2$, поэтому полная система вычетов по модулю q содержит $|q|^2$ элементов. Значит, $|q|^2 \leqslant 4k+1$. Следовательно, если $4k+1 < |m|^2$, то l не делится на m. А так как $mH_k(m)$ делится на l, то $H_k(m)$ делится на l и c_k делится на m. Таким образом, если m — нечетное простое число, то c_1, $c_2, \ldots, c_{\frac{p-5}{4}}$, где $p = |m|^2$, делятся на m. Из соотношения

$$A_{\frac{p-1}{4}} + A_{\frac{p-5}{4}}\varphi^4 + \ldots + \varphi^{p-1} = \\ = (1 + A_1\varphi^4 + \ldots + A_{\frac{p-1}{4}}\varphi^{p-1})(c_0 + c_1\varphi^4 + c_2\varphi^8 + \ldots)$$

следуют равенства

$$A_{\frac{p-1}{4}} = c_0,$$
$$A_{\frac{p-5}{4}} = c_0 A_1 + c_1,$$
$$\dots\dots\dots\dots\dots$$
$$A_1 = c_0 A_{\frac{p-1}{4}} + c_1 A_{\frac{p-5}{4}} + \dots + c_{\frac{p-5}{4}}.$$

Поэтому числа A_1, A_2, ..., $A_{\frac{p-1}{4}} = m$ делятся на m. □

Задачи

1. Докажите, что если a, b, x, $y \in \mathbb{R}$ и $x + iy = \sqrt{a + ib}$, то x и y выражаются через a и b с помощью квадратных радикалов.

2. (Теорема Эйзенштейна) Пусть $f(x) = a_0 x^n + a_1 x^{n-1} + \dots + a_n$, где $a_j \in \mathbb{Z}[i]$, причем для некоторого простого над $\mathbb{Z}[i]$ числа m коэффициенты $a_1, \dots, a_n$ делятся на m, a_n не делится на m^2 и a_0 не делится на m. Докажите, что многочлен f неприводим над $\mathbb{Z}[i]$.

3. Докажите, что если m — нечетное простое над $\mathbb{Z}[i]$ число, то многочлен $\varphi^{p-1} + A_1 \varphi^{p-5} + \dots + A_{\frac{p-1}{4}}$ из теоремы 1 неприводим.

§ 4. Доказательство теоремы Абеля о делении лемнискаты

В этом параграфе мы приведем два доказательства теоремы Абеля. Одно из них — классическое — принадлежит Эйзенштейну, другое — более современное — Роузену [B14]. Краеугольным камнем этих двух доказательств, равно как и оригинального доказательства самого Абеля, служит тот факт, что решетка периодов функции $\varphi(\alpha)$ инвариантна относительно умножения на комплексную единицу i. Лучше всего это видно в доказательстве Роузена, поэтому с него мы и начнем.

В подходе Роузена вместо лемнискатической функции $\varphi(\alpha)$ используется функция Вейерштрасса $\wp(z)$, отвечающая решетке

$$\Lambda = \{2a\omega + 2bi\omega \mid a, b \in \mathbb{Z}\}.$$

Отметим, что эта решетка содержится в решетке периодов функции φ, но не совпадает с ней. В конце этого параграфа мы покажем, что для указанной решетки $g_2 = -1/4$ и $g_3 = 0$, т. е.

$$\wp'^2(z) = 4\wp^3(z) - \frac{1}{4}\wp(z).$$

Возможность перехода к функции $\wp(z)$ связана с тем, что справедливо следующее утверждение.

Лемма 1. *Если отрезок длиной $\wp(\alpha)$ можно построить с помощью циркуля и линейки, то отрезок длиной $\varphi(\alpha)$ тоже можно построить.*

Доказательство. По модулю решетки Λ нули функции φ имеют вид 0, ω, $i\omega$, $(1+i)\omega$, а ее полюсы имеют вид

$$\frac{(1+i)\omega}{2}, \frac{(3+i)\omega}{2}, \frac{(1+3i)\omega}{2}, \frac{(3+3i)\omega}{2}.$$

Функция $\wp'(z)$ тоже имеет нули ω, $i\omega$, $(1+i)\omega$, а точка 0 — полюс этой функции. Кроме того,

$$\wp\left(\frac{1+i}{2}\omega\right) = \wp\left(\frac{3+3i}{2}\omega\right) \quad \text{и} \quad \wp\left(\frac{3+i}{2}\omega\right) = \wp\left(\frac{1+3i}{2}\omega\right).$$

Рассмотрим функцию

$$g(z) = \frac{\wp'(z)}{(\wp(z) - \wp(\frac{1+i}{2}\omega))(\wp(z) - \wp(\frac{3+i}{2}\omega))}.$$

В окрестности нуля $\wp(z) = z^{-2} + \dots$ и $\wp'(z) = -2z^{-3} + \dots$, поэтому $g(0) = 0$. Следовательно, функция g имеет те же самые нули и полюсы, что и функция φ. Поэтому $\varphi(z) = Cg(z)$.

Докажем теперь, что отрезки длиной $\wp\left(\frac{\omega}{2}\right)$, $\wp\left(\frac{1+i}{2}\omega\right)$ и $\wp\left(\frac{3+i}{2}\omega\right)$ можно построить. Заметим, прежде всего, что отрезки длиной $\wp(\omega)$, $\wp(i\omega)$ и $\wp((1+i)\omega)$ можно построить, так как эти числа являются корнями уравнения $4x^3 - \frac{1}{4}x = 0$. Кроме того, легко проверить, что

$$\wp(z \pm iz) = \mp\frac{i}{32}\,\frac{16\wp^2(z) - 1}{\wp(z)}.$$

Пусть задано $\wp(\alpha)$. Рассмотрим $x = \dfrac{\alpha}{1+i}$ и $y = \dfrac{x}{1-i} = \dfrac{\alpha}{2}$. Чтобы найти $\wp\left(\dfrac{\alpha}{2}\right)$, достаточно решить квадратные уравнения

$$\wp(\alpha) = -\frac{i}{32}\,\frac{16\wp^2(x)-1}{\wp(x)} \quad \text{и} \quad \wp(x) = \frac{i}{32}\,\frac{16\wp^2(y)-1}{\wp(y)}.$$

Если можно построить $\wp(\alpha)$, то можно построить и $\wp'(\alpha) = \sqrt{4\wp^3(\alpha) - \dfrac{1}{4}}$. Следовательно, $g(\omega/2)$ можно построить, а так как $\varphi(\omega/2) = 1$, то можно построить и константу C. В итоге получаем, что если можно построить $\wp(\alpha)$, то можно построить $g(\alpha)$, а значит, можно построить и $\varphi(\alpha)$. □

Таким образом, для доказательства теоремы Абеля нужно проверить, что если $n = 2^a p_1 \dots p_m$, где p_i — различные простые числа Ферма, то отрезки длиной $\wp\left(\dfrac{k\omega}{n}\right)$, где $k = 1, \dots, n-1$, можно построить.

Отображение $z \longmapsto (\wp(z), \wp'(z))$ можно рассматривать как гомеоморфизм тора $\mathbb{C}/\Lambda$ на кривую E, заданную уравнением $y^2 = 4x^3 - \dfrac{1}{4}x$. При этом сложение точек тора индуцирует сложение точек кривой E. Элементы группы E, порядок которых делит n, образуют подгруппу

$$E_n = \left\{ \left(\wp\left(\frac{2a\omega + 2bi\omega}{n}\right),\ \wp'\left(\frac{2a\omega + 2bi\omega}{n}\right)\right) \mid 0 \leqslant a, b < n \right\}.$$

Нулевой элемент соответствует $a = b = 0$. Это — бесконечно удаленная точка.

Группа E_n аналогична группе C_n для окружности. Продолжая аналогию, покажем, что если (a,b) и (c,d) — точки E, то

$$(a,b) + (c,d) = (f(a,b,c,d), g(a,b,c,d)),$$

причем $\sigma f(u) = f(\sigma u)$ и $\sigma g(u) = g(\sigma u)$ для любого автоморфизма σ поля $\mathbb{C}$. Теорема сложения

$$\wp(z_1+z_2) = -\wp(z_1) - \wp(z_2) + \frac{1}{4}\left(\frac{\wp'(z_1)-\wp'(z_2)}{\wp(z_1)-\wp(z_2)}\right)^2 \qquad (4.1)$$

показывает, что

$$f(a,b,c,d) = -a - c - \frac{1}{4}\left(\frac{b-d}{a-c}\right)^2$$

при $a \neq c$. Дифференцируя равенство (4.1) и учитывая, что

$$\wp''(z) = 6\wp^2(z) - \frac{1}{2}g_2 = 6\wp^2(z) - \frac{1}{8},$$

можно представить g в виде рациональной функции от a, b, c, d с рациональными коэффициентами. В случае, когда $z_1 \equiv z_2 \pmod{\Lambda}$, можно воспользоваться формулой

$$\wp(2z) = -2\wp(z) + \frac{1}{4}\left(\frac{\wp''(z)}{\wp'(z)}\right)^2.$$

Если же $z_1 \equiv -z_2 \pmod{\Lambda}$, то остается справедливой формула (4.1), нужно лишь считать выражения в обеих частях бесконечными.

С помощью функций f и g можно получить функции f_n и g_n, для которых

$$(f_n(x,y), g_n(x,y)) = n(x,y).$$

Точки E_n задаются при этом уравнениями $f_n(x,y) = \infty$ и $g_n(x,y) = \infty$. Для точек, отличных от бесконечно удаленной точки, это означает, что обращаются в нуль знаменатели дробей f_n и g_n.

Теперь можно рассмотреть поле K_n, порожденное над $\mathbb{Q}$ координатами конечных точек E_n. Повторив для E_n такие же рассуждения, как для C_n, получим, что группа G_n автоморфизмов поля K_n над $\mathbb{Q}$ изоморфна некоторой подгруппе группы $\mathrm{Aut}(E_n)$. Так как

$$E_n \approx \left(\frac{1}{n}\Lambda\right)/\Lambda \approx \Lambda/n\Lambda \approx \mathbb{Z}/n\mathbb{Z} \oplus \mathbb{Z}/n\mathbb{Z},$$

то $\mathrm{Aut}(E_n) \approx GL_2(\mathbb{Z}/n\mathbb{Z})$. В случае простого n порядок группы $GL_2(\mathbb{Z}/n\mathbb{Z})$ равен количеству базисов пространства $(\mathbb{Z}/n\mathbb{Z})^2$, т. е. равен $(n^2-1)(n^2-n)$. Это число делится на $n(n+1)$, поэтому ни при каком $n \geqslant 2$ оно не может быть степенью двойки. Наши рассуждения зашли в тупик!

Выручит нас свойство, которым мы неоднократно пользовались при изучении многочленов деления лемнискаты, а именно, инвариантность решетки периодов относительно умножения на i.

В данном случае это означает, что $\wp(iz)$ и $\wp'(iz)$ выражаются через $\wp(z)$ и $\wp'(z)$ по формулам $\wp(iz) = -\wp(z)$ и $\wp'(iz) = i\wp'(z)$. В самом деле,

$$\wp(z) = z^{-2} + {\sum_{\lambda\in\Lambda}}'((z-\lambda)^{-2} - \lambda^{-2}).$$

Из инвариантности решетки Λ относительно умножения на i следует, что $\wp(iz) = -\wp(z)$. Дифференцируя это равенство, получаем $i\wp'(iz) = -\wp'(z)$. Таким образом, действие i на торе $\mathbb{C}/\Lambda$ индуцирует действие i на кривой E, которое задается формулой $i(x,y) = (-x, iy)$. На группе $\Lambda/n\Lambda$, изоморфной E_n, действие элемента $k + il \in \mathbb{Z}[\,i\,]$ задается формулой

$$(2a\omega + 2bi\omega) \,(\mathrm{mod}\, n) \longmapsto (k+il)\,(2a\omega + 2bi\omega) \,(\mathrm{mod}\, n).$$

Это действие переносится на группу E_n.

Пусть $F = \mathbb{Q}(i)$, F_n — поле, порожденное координатами точек E_n над F, G_n — группа автоморфизмов поля F_n над F. Если $\sigma \in G_n$, то $\sigma(i) = i$, поэтому $\sigma(i(x,y)) = (-\sigma x, i\sigma y) = i\sigma(x,y)$. Кроме того $\sigma((a,b)+(c,d)) = \sigma((a,b)) + \sigma((c,d))$. Следовательно, G_n — подгруппа группы автоморфизмов $\mathbb{Z}[\,i\,]$ — модуля $\Lambda/n\Lambda$.

Для отображения $a+ib \longmapsto (k+il)\,(a+ib)$, где a и b берутся по модулю n, обратным будет отображение

$$a + ib \longmapsto \frac{k-il}{k^2+l^2}\,(a+ib).$$

Оно определено тогда и только тогда, когда число k^2+l^2 взаимно просто с n. Для получения попарно различных отображений нужно считать, что $0 \leqslant k,\ l \leqslant n-1$. Таким образом, порядок группы автоморфизмов $\mathbb{Z}[\,i\,]$ — модуля $\Lambda/n\Lambda$ равен количеству пар чисел (k,l), где $0 \leqslant k,\ l \leqslant n-1$ и k^2+l^2 взаимно просто с n. Обозначим их количество $\Phi(n)$. Вычисление $\Phi(n)$ разобьем на леммы.

Лемма 2. *Если числа p и q взаимно просты, то* $\Phi(pq) = \Phi(p)\Phi(q)$.

Доказательство. Пусть $0 \leqslant a_1,\ b_1 \leqslant p-1$ и $0 \leqslant a_2,\ b_2 \leqslant q-1$. Тогда пары $(a,b) = (a_1q + a_2p, b_1q + b_2p)$ образуют полную систему пар вычетов по модулю pq. Кроме того,

$$a^2 + b^2 = (a_1^2 + b_1^2)\,q^2 + (a_2^2 + b_2^2)\,p^2.$$

Поэтому взаимная простота чисел a^2+b^2 и pq эквивалентна взаимной простоте чисел $a_1^2+b_1^2$ и p и чисел $a_2^2+b_2^2$ и q. □

Лемма 3. *Если p — простое число вида $4k+3$, то* $\Phi(p) = p^2 - 1$.

Доказательство. Нужно доказать, что если a^2+b^2 делится на p, то оба числа a^2 и b^2 делятся на p. Предположим, что a^2+b^2 делится на p, но хотя бы одно из чисел a и b не делится на p. Тогда оба числа a и b не делятся на p, поэтому согласно малой теореме Ферма $a^{p-1} \equiv 1\,(\mathrm{mod}\, p)$ и $b^{p-1} \equiv 1\,(\mathrm{mod}\, p)$, а значит, $a^{p-1} + b^{p-1} \equiv 2\,(\mathrm{mod}\, p)$. С другой стороны, число

$$a^{p-1} + b^{p-1} = a^{4k+2} + b^{4k+2} = (a^2)^{2k+1} + (b^2)^{2k+1}$$

делится на a^2+b^2, поэтому оно делится на p. □

Лемма 4. *Если p — простое число вида $4k+1$, то* $\Phi(p) = (p-1)^2$.

Доказательство. Прежде всего докажем, что любое простое число p вида $4k+1$ представимо в виде суммы двух квадратов. Наиболее простое доказательство этого утверждения предложено Д. Цагиром. Рассмотрим множество всех решений уравнения $x^2 + 4yz = p$ в натуральных числах. Достаточно доказать, что у этого уравнения есть решение, для которого выполнено равенство $y = z$.

Это равносильно тому, что инволюция $\sigma(x,y,z) = (x,z,y)$, определенная на множестве решений, имеет неподвижную точку. Напомним, что *инволюцией* называют такое отображение f, что $f(f(x)) = x$ для всех x. Если $f(x_0) = x_0$, то x_0 называют *неподвижной точкой*. На множестве, состоящем из нечетного числа элементов, любая инволюция имеет неподвижную точку. Поэтому достаточно доказать, что общее количество решений данного уравнения нечетно.

Построим для этого инволюцию τ, имеющую ровно одну неподвижную точку. Определим

$$\tau(x,y,z) = \begin{cases} (x+2z, z, y-x-z) & \text{при} \quad x < y-z, & (A)\\ (2y-x, y, x-y+z) & \text{при} \quad y-z < x < 2y, & (B)\\ (x-2y, x-y+z, y) & \text{при} \quad 2y < x. & (C) \end{cases}$$

Легко проверить, что $x \neq 2y$ и $x \neq y-z$. Кроме того, любое решение действительно переводится отображением τ в решение. Разобьем решения на три типа в соответствии с тем, какое именно из неравенств выполняется: $x < y-z$, $y-z < x < 2y$ или

$2y < x$. Отображение τ переводит тип A в тип C, тип B в тип B, а тип C в тип A. Теперь легко проверить, что τ — инволюция.

Неподвижной может быть лишь точка типа B. Из равенства

$$(x, y, z) = (2y - x, y, x - y + z)$$

следует, что $y = x$. Поэтому $p = x(x + 4z)$, т. е. $x = y = 1$ в силу простоты числа p. Следовательно, неподвижная точка ровно одна, а именно $(1, 1, k)$, т. к. $p = 4k + 1$.

Докажем теперь, что при фиксированном $a \neq 0$ уравнение $x^2 + a^2 \equiv 0 \pmod p$ имеет ровно два решения. В самом деле, существуют такие ненулевые числа b и c, что $b^2 + c^2 \equiv 0 \pmod p$. Домножив это равенство на $(ac^{-1})^2$, получим $b_1^2 + a^2 \equiv 0 \pmod p$, где $b_1 = abc^{-1}$. Следовательно, $x^2 \equiv b_1^2 \pmod p$ это уравнение имеет решения $x = \pm b_1$. Таким образом, в $\Phi(p)$ не входят лишь $2(p-1)$ пар с ненулевыми a и b и пара $(0, 0)$. Поэтому $\Phi(p) = p^2 - 1 - 2(p-1) = (p-1)^2$. □

Очевидно также, что $\Phi(2) = 2$.

Лемма 5. *Пусть p — простое число, $k \geqslant 1$. Тогда* $\Phi(p^k) = (p^{k-1})^2 \Phi(p)$.

Доказательство. Числа $a + a_1 p$, где $0 \leqslant a \leqslant p-1$, $0 \leqslant a_1 \leqslant p^{k-1} - 1$ образуют полную систему вычетов по модулю p^k. Число $(a + a_1 p)^2 + (b + b_1 p)^2$ не взаимно просто с p^k тогда и только тогда, когда оно не взаимно просто с p, т. е. $a^2 + b^2 \equiv 0 \pmod p$. Остается заметить, что каждой паре (a, b), входящей в $\Phi(p)$, соответствует $(p^{k-1})^2$ пар, входящих в $\Phi(p^k)$. □

Теперь уже легко проверить, что $\Phi(n)$ — степень двойки тогда и только тогда, когда $n = 2^a p_1 \dots p_m$, где p_i — различные простые числа Ферма. В самом деле, $\Phi(2^a p_1^{k_1} \dots p_m^{k_m})$ может быть степенью двойки лишь в том случае, когда $k_1 = \dots = k_m = 1$. Если $p = 4k + 1$, то $\Phi(p) = (p-1)^2$. Это число — степень двойки лишь в том случае, когда $p = 1 + 2^a$. Пусть $p = 4k + 3$ и $\Phi(p) = p^2 - 1 = (p-1)(p+1)$ есть степень двойки. Последовательные четные числа $p - 1$ и $p + 1$ могут быть степенями двойки лишь в том случае, когда $p = 3$.

Для завершения доказательства теоремы Абеля нам остается лишь проверить, что для рассматриваемой решетки выполнены равенства

$$g_2 = 60 \sum{}' (2a\omega + 2bi\omega)^{-4} = \frac{1}{4}$$

и

$$g_3 = 140 \sum{}' (2a\omega + 2bi\omega)^{-6} = 0.$$

Второе равенство очевидно, так как $g_3(\Lambda) = -g_3(i\Lambda)$, а в нашем случае $i\Lambda = \Lambda$. Основная трудность заключается в доказательстве равенства $\sum{}' (a\omega + bi\omega)^{-4} = \frac{1}{15}$.

Рассмотрим три решетки:

$$L_0 = \{a\omega + bi\omega\},$$
$$L_1 = \left\{ \frac{a\omega + bi\omega}{2} \;\middle|\; a \text{ и } b \text{ нечетны} \right\},$$
$$L_2 = \left\{ \frac{a\omega + bi\omega}{2} \;\middle|\; a - b \text{ нечетно} \right\}.$$

Тогда $\frac{1}{2} L_0 = L_0 \cup L_1 \cup L_2$ и $L_2 = \frac{1+i}{2} L_1$. Обозначим $\sum_{l \in L}{}' l^{-4}$ через $|L|$. Тогда

$$16\,|L_0| = \left|\frac{1}{2} L_0\right| = |L_0| + |L_1| + |L_2|$$

и

$$|L_2| = \left(\frac{2}{1+i}\right)^4 |L_1| = -4\,|L_1|,$$

поэтому $|L_1| = -5\,|L_0|$. Чтобы доказать требуемое равенство $|L_0| = 1/15$, получим еще одно соотношение между $|L_1|$ и $|L_0|$.

Воспользуемся для этого тем, что L_0 — решетка нулей функции $\varphi(z)$, а L_1 — решетка ее полюсов. Учитывая, что $\varphi'(0) = 1$, получим

$$\varphi(z) = z \prod_{\alpha \in L_0}{}' \left(1 - \frac{z}{\alpha}\right) \prod_{\beta \in L_1}{}' \left(1 - \frac{z}{\beta}\right)^{-1},$$

причем бесконечные произведения нужно понимать как пределы при $N \to \infty$ конечных произведений, определяемых условиями $|\alpha|, |\beta| \leqslant N$. Ненулевые элементы решеток L_0 и L_1 можно разбить на четверки вида $\{\pm\gamma, \pm i\gamma\}$, поэтому

$$\varphi(z) = z \prod{}' \left(1 - \frac{z^4}{\alpha^4}\right) \prod{}' \left(1 - \frac{z^4}{\beta^4}\right)^{-1},$$

где $0 \leqslant \arg\alpha, \arg\beta < \pi/2$. Следовательно,

$$z\frac{\varphi'(z)}{\varphi(z)} = z\frac{d}{dz}\ln\varphi(z) = 1 + (|L_1| - |L_0|)\,z^4 + \ldots \qquad (4.2)$$

Функция $z^{-1}\varphi(z)$ не изменяется при замене z на $-z$ или $\pm iz$, поэтому $\varphi(z) = z(1 + cz^4 + \ldots)$. Кроме того, $(\varphi'(z))^2 = 1 - \varphi^4(z)$, поэтому $(1 + 5cz^4 + \ldots)^2 = 1 - z^4(1 + cz^4 + \ldots)^4$, а значит, $c = -1/10$. Следовательно,

$$z\frac{\varphi'(z)}{\varphi(z)} = 1 + 4cz^4 + \ldots = 1 - \frac{2}{5}z^4 + \ldots \qquad (4.3)$$

Сравнив выражения (4.2) и (4.3), получим $|L_1| - |L_0| = -2/5$. А так как $|L_1| = -5|L_0|$, то $|L_0| = 1/15$.

Приведем теперь вкратце доказательство Эйзенштейна, точнее, его модифицированную версию, предложенную Мельниковым [В12]. Это доказательство основано на свойствах многочленов деления лемнискаты, которые были получены в § 3.

Итак, рассмотрим числа $\varphi\left(\frac{k\Omega}{n}\right)$, $k = 1, \ldots, n-1$, где n — некоторое натуральное число. В силу теоремы сложения достаточно ограничиться случаем, когда n — простое число. Для $n = 2$ теорема была доказана в § 3, где мы нашли явный вид уравнения деления и делим его в квадратичных радикалах. Поэтому предположим, что n — нечетное простое число. Если $n \equiv 3 \pmod 4$, то оно остается простым и в кольце $\mathbb{Z}[i]$, если же $n \equiv 1 \pmod 4$, то оно распадается в кольце $\mathbb{Z}[i]$ в произведение двух простых сопряженных множителей $m = a + bi$ и $m = a - bi$. Вычислив $\varphi\left(\frac{k\Omega}{n}\right)$ и $\varphi\left(\frac{k\Omega'}{m}\right)$, по теореме сложения найдем $\varphi\left(\frac{k\Omega}{n}\right)$. Таким образом, будем считать, что n — простое число в $\mathbb{Z}[i]$.

Обратимся теперь к уравнению деления лемнискаты

$$\Phi(x) = x^{p-1} + A_1x^{p-5} + A_2x^{p-5} + \ldots + A_{\frac{p-1}{4}} = 0.$$

Согласно теореме 3 § 2 для простого над $\mathbb{Z}[i]$ числа n все коэффициенты A_l делятся на n и, кроме того, $A_{\frac{p-1}{4}} = n$. Применяя критерий Эйзенштейна, мы получаем, что многочлен $\Phi(x)$ неприводим для простого n.

Пусть g — примитивный корень по модулю простого числа n. Очевидно, что g можно выбрать нечетным. Обозначим корни уравнения (4.4) через

$$x_0 = \varphi\left(\frac{\Omega}{n}\right), \quad x_1 = \varphi\left(g\frac{\Omega}{n}\right), \ldots, \quad x_{p-2} = \varphi\left(g^{p-2}\frac{\Omega}{n}\right),$$

получим, что каждый из них будет одной и той же рациональной функцией предыдущего, т. е.

$$x_{l+1} = x_l\frac{\Phi_g(x_l)}{\Psi_g(x_l)}.$$

Это означает, что уравнение деления лемнискаты абелево и поэтому разрешимо в радикалах (исчерпывающее изложение теории абелевых уравнений можно найти в двухтомнике Бернсайда и Пантона [Б6]). Все эти радикалы будут квадратичными тогда и только тогда, когда $p = 2^\alpha + 1$. Если n — простое гауссово число, то p также простое и имеет вид $p = 2^{2^\beta} + 1$. Если же p — квадрат простого числа $q \equiv 3 \pmod 4$, то равенство $q = 2^\alpha + 1$ возможно лишь при $q = 3$.

§ 5. Несколько замечаний о кривых Серре

«„Wenn die Könige bauen, haben die Kärrer zu tun“ *) — сказал немецкий поэт... Но возчикам нужны дороги. В истории нашей науки нередко случалось так, что король открывал новый путь в землю обетованную, а его наследники, предпочитая свои тропинки, оставляли этот путь зарастать чертополохом.»

Эта сентенция Андрэ Вейля (цит. по [Б12]) вполне применима к кривым Серре. Достаточно сказать, что даже в знаменитых «Лекциях о развитии математики в XIX столетии» Феликса Клейна имя Серре встречается всего лишь один раз, да и то совсем по иному поводу. Единственное упоминание об этих кривых, которое нам удалось найти в математической литературе, принадлежит Джорджу Сальмону [Б24] и относится к викторианским временам...

*) «Когда строят короли, работу получают возчики» (*перев. с немецкого*). Строка из стихотворения «Кант и его последователи», приписываемого Шиллеру.

Около двух лет тому назад кривые Серре привлекли внимание югославского математика А. Липковского. Ниже мы приводим некоторые его результаты, относящиеся к этому предмету.

Для исследования кривых Серре нам понадобятся некоторые свойства плоских алгебраических кривых. Перечислим их вкратце, отсылая за подробностями к руководствам [Б29, Б14].

Будем рассматривать в комплексной проективной плоскости определенную проективную систему координат. Пусть $f(x, y, z)$ — неприводимый однородный многочлен степени n от неизвестных x, y, z с комплексными коэффициентами. Если некоторая тройка (a_1, a_2, a_3) координат точки P удовлетворяет уравнению $f(x, y, z) = 0$, то все другие значения координат этой точки также удовлетворяют тому же уравнению, поскольку $f(\rho a_1, \rho a_2, \rho a_3) = \rho^n f(a_1, a_2, a_3)$. Множество всех точек P с указанным свойством называется *неприводимой алгебраической кривой* степени n.

Пусть C — некоторая кривая, задаваемая уравнением $f(x, y, z) = 0$. Так как многочлен $f(x, y, z)$ неприводим, то он, в частности, не делится на z, и поэтому существует соответствующий неоднородный многочлен $f(x, y, 1)$. Обозначим его через $F(x, y)$. Уравнение $F(x, y) = 0$ называется уравнением кривой C в соответствующей аффинной координатной системе. Ясно, что решением уравнения $F(x, y) = 0$ будут те точки кривой C, которые не лежат на бесконечно удаленной прямой $z = 0$.

Пусть $A = (a_1, a_2, a_3)$ и $B = (b_1, b_2, b_3)$ — две различные точки, лежащие на кривой C. Точка $P = (x_1, x_2, x_3)$ лежит на прямой L, соединяющей точки A и B, тогда и только тогда, когда $x_k = a_k s + b_k t$ при некоторых s и t. Значения s и t, дающие точки, лежащие на кривой C, будут решениями уравнения

$$f(as + bt) = f(a_1 s + b_1 t, a_2 s + b_2 t, a_3 s + b_3 t) = 0.$$

Так как f — неприводимый многочлен, то $f(as + bt)$ не является нулем тождественно по s и t. Следовательно, он будет однородным многочленом степени n от этих переменных и уравнение $f(as + bt) = 0$ удовлетворяется точно n значениями отношения $s : t$, если считать корни этого уравнения в соответствии с их кратностями.

Каждое значение отношения $s : t$ определяет точку, общую для L и C. Точку, отвечающую r-кратному корню, удобно считать r-кратной точкой пересечения L и C.

Рассмотрим теперь более внимательно пересечение кривой C и прямой L в том случае, когда L проходит через данную точку P кривой C. Пусть $f(x, y) = 0$ — уравнение кривой C в аффинной системе координат, в которой P имеет координаты (a, b). Параметрические уравнения прямой L имеют вид:

$$x = a + \lambda t, \quad y = b + \mu t.$$

Прямая L определяется отношением $\lambda : \mu$. Точки пересечения L и C соответствуют корням уравнения

$$f(a + \lambda t, b + \mu t) = 0.$$

Разложим левую часть этого уравнения по степеням t. Учитывая, что $f(a, b) = 0$, получаем

$$t(f_x \lambda + f_y \mu) + \frac{t^2}{2!}(f_{xx}\lambda^2 + 2f_{xy}\lambda\mu + f_{yy}\mu^2) + \ldots = 0,$$

где $f_x, f_y, \ldots$ — значения производных f в точке P. Возможны два случая.

а) Предположим, что f_x и f_y не равны 0 одновременно. Тогда почти каждая прямая, проходящая через P имеет с кривой C однократное пересечение в точке P. Единственным исключением будет прямая, соответствующая значению $\lambda : \mu$, для которого $f_x \lambda + f_y \mu = 0$. Эта прямая является касательной к C в точке P.

б) Предположим теперь, что все производные f до порядка $(r - 1)$ включительно $(r > 1)$ равны 0 в точке P и хотя бы одна производная порядка r отлична от 0 в этой точке. Тогда каждая прямая, проходящая через P, имеет с кривой C по меньшей мере r-кратное пересечение в точке P. При этом точно r прямых имеют более, чем r-кратное пересечение с кривой C. Эти исключительные прямые, касательные к кривой в точке P соответствуют решениям уравнения

$$f_{x^r}\lambda^r + C_r^1 f_{x^{r-1}y}\lambda^{r-1}\mu + \ldots + f_{y^r}\mu^r = 0$$

и должны при подсчете их числа браться такое же число раз, какова кратность соответствующего корня. В этом случае точка P называется *r-кратной точкой* (*точкой кратности r*) кривой C.

Точка кривой C, имеющая кратность 1, называется *простой* точкой этой кривой. Точки кратности 2 и более называются *особыми* точками. Точка кратности r называется *обыкновенной*

r-кратной точкой, если в ней существует r различных касательных. Необходимое и достаточное условие того, что точка (a,b,c) является особой, дается равенствами

$$f(a,b,c)=\frac{\partial f(a,b,c)}{\partial x}=\frac{\partial f(a,b,c)}{\partial y}=\frac{\partial f(a,b,c)}{\partial z}=0$$

в проективных координатах и соответственно равенствами

$$F(a,b)=\frac{\partial F(a,b)}{\partial x}=\frac{\partial F(a,b)}{\partial y}=0$$

в аффинных координатах.

Критерий для нахождения особых точек в проективных координатах выражается следующим образом. Точка P тогда и только тогда является r-кратной точкой кривой $f(x,y,z)=0$, когда в этой точке все производные от f порядка $r-1$ обращаются в нуль, но существует производная порядка r, отличная от нуля.

Рассмотрим две проективные плоскости $\Pi_1(x_1,x_2,x_3)$ и $\Pi_2(y_1,y_2,y_3)$ и зададим отображение T: $\Pi_1\to\Pi_2$, полагая

$$y_i=x_jx_k,$$

где $i,j,k=1,2,3$, причем числа i,j,k различны. Отображение T называется *квадратичным преобразованием* или *раздутием* плоскости Π_1 в Π_2. Оно обладает следующими свойствами.

1) Каждая точка плоскости Π_1, за исключением точек $(1,0,0)$, $(0,1,0)$ и $(0,0,1)$, отображается в определенную точку плоскости Π_2. Три исключительные точки называются *фундаментальными точками преобразования* T. Их образы не определены.

(2) Все точки прямой $x_i=0$, кроме фундаментальных, отображаются в точку $y_i=1$, $y_j=0$, $y_k=0$. Три прямые $x_i=0$ называются *иррегулярными прямыми* отображения.

Обозначим через T' отображение

$$x_i=y_jy_k$$

плоскости Π_2 в Π_1. Оно, очевидно, также обладает свойствами, подобными свойствам 1) и 2). Кроме того, имеет место следующее свойство.

3) Если $(x)=(x_1,x_2,x_3)$ — точка, не лежащая на иррегулярных прямых отображения T, то точка $(y)=T(x)$ не лежит на иррегулярных прямых отображения T', причем $T'(y)=(x)$.

Поскольку аналогичное свойство имеет место также и для T', мы видим, что T и T' определяют взаимно однозначное соответствие между точками плоскостей Π_1 и Π_2, не лежащими на иррегулярных прямых.

Если

$$f(x_1,x_2,x_3)=0$$

— кривая на плоскости Π_1, то образы ее точек при отображении T будут удовлетворять уравнению

$$g(y)=f(y_2y_3,y_1y_3,y_1y_2)=0.$$

Кривая g называется *алгебраическим образом* кривой f при отображении T. Пусть f — неприводимый многочлен и пусть g — алгебраический образ f. Если $g(y)=\pi(y)f'(y)$, где $\pi(y)$ — произведение степеней y_i, а многочлен f' не делится ни на один из y_i, то кривую $f'=0$ называют *образом* кривой f при отображении T. Имеет место следующая фундаментальная теорема.

Теорема 1. *Любую неприводимую кривую последовательными раздутиями можно преобразовать в неприводимую кривую, имеющую лишь обыкновенные особенности.*

Для исследования структуры особых точек кривой удобно параметризовать части этой кривой формальными степенными рядами. Обозначим через $\mathbb{C}[[t]]$ поле формальных степенных рядов от переменной t с комплексными коэффициентами. Каждый отличный от нуля элемент $\mathbb{C}[[t]]$ может быть однозначно записан в виде

$$u=t^{-k}(a_0+a_1t+\ldots),$$

где k — целое число и $a_0\neq 0$. Число k называется порядком ряда u и будет обозначаться через $O(u)$.

Пусть $f(x_1,x_2,x_3)=0$ — уравнение алгебраической кривой C в комплексной проективной плоскости. Элементы u_1,u_2,u_3 поля $\mathbb{C}[[t]]$ называются координатами некоторой *параметризации* кривой C, если

1) $f(u_1,u_2,u_3)=0$.

2) Не существует ненулевого элемента $e \in \mathbb{C}[[t]]$, для которого $eu_i \in \mathbb{C}$, $i = 1, 2, 3$.

Пусть $(u) = (u_1, u_2, u_3)$ — произвольная параметризация кривой C и $h = -\min O(u_i)$. Тогда, если положить $v_i = t^h u_i$, то элементы v_i будут определять ту же самую параметризацию. При этом $v_i \in \mathbb{C}[[t]]$ и хотя бы одно из значений $O(v_i)$ равно нулю. Следовательно, хотя бы одно из значений $v_i(0) = a_i \neq \neq 0$. Точка $(a) = (a_1, a_2, a_3)$ называется в таком случае *центром* параметризации. Очевидно, что при преобразовании координат координаты центра будут вести себя подобно координатам точек. Отсюда следует, что параметризация имеет однозначно определенный центр.

Если $(u) = (u_1, u_2, u_3)$ — некоторая параметризация и если для ненулевого элемента $s \in \mathbb{C}[[t]]$ порядок $O(s) > 0$, то совокупность $(v) = (u_1(s), u_2(s), u_3(s))$ также будет параметризацией с тем же центром. В случае, когда $O(s) = 1$, две такие параметризации называются *эквивалентными*.

Предположим, что $u_i \in \mathbb{C}[[t^r]]$ для некоторого $r > 1$. Тогда можно упростить степенные ряды, заменяя t^r новой неизвестной t'. В этом случае параметризация $(u) = (u_1, u_2, u_3)$, а также любая эквивалентная ей, называется *приводимой*. Если такого упрощения не существует, параметризация называется *неприводимой*. Класс эквивалентности неприводимых параметризаций кривой C называется *ветвью* этой кривой. Общий центр всех параметризаций называется центром ветви. Можно показать, что любая точка кривой C является центром хотя бы одной ветви.

Если $(u) = (u_1, u_2, u_3)$ — параметризация ветви P, для которой все $O(u_i) \geqslant 0$ и хотя бы одно из значений $O(u_i) = 0$, и если $g(x_1, x_2, x_3)$ — произвольный однородный многочлен, то мы можем определить порядок $O_P(g)$ многочлена g, как порядок степенного ряда $g(u_1, u_2, u_3)$. Положительное число $r = \min O_P(L)$, где L — любая прямая, проходящая через центр ветви P, называется *порядком ветви*.

Одно из наиболее важных применений понятия ветви и ее порядка содержится в следующей теореме.

Теорема 2. 1) *Если Q — некоторая r-кратная точка кривой C, то сумма порядков ветвей C, имеющих Q своим центром, равна r.*

2) *Точка кривой C будет простой тогда и только тогда, когда она является центром единственной ветви.*

Обратимся теперь к кривым Серре. Напомним их определение. Пусть p — фиксированное рациональное число. Рассмотрим треугольник OPV с вершиной O, находящейся в начале координат, сторонами OP и PM, равными соответственно $\sqrt{p}$ и $\sqrt{p+1}$, и углами при вершинах O и M, равными α и β (рис. 29). Будем изменять треугольник OPM так, чтобы точка O оставалась неподвижной, длины сторон OP и PM оставались постоянными, а угол ω между осью Ox и стороной OM определялся бы следующим соотношением:

$$\cos\omega = \cos(p\alpha - (p+1)\beta). \tag{5.1}$$

Тогда геометрическое место точек вершины M называется кривой Серре, отвечающей параметру p, и обозначается через S_p.

Пусть ρ — длина отрезка OM. Тогда координаты точки M имеют вид $x = \rho\cos\omega$, $y = \rho\sin\omega$. Из теоремы косинусов вытекает, что

$$\cos\alpha = \frac{\rho^2 - 1}{2\rho\sqrt{p}}, \quad \cos\beta = \frac{\rho^2 + 1}{2\rho\sqrt{p+1}}, \tag{5.2}$$

поэтому для рационального числа p выражение (5.1) можно представить в виде многочлена от $\cos\alpha$, $\cos\beta$, $\sin\alpha$ и $\sin\beta$. Другими словами, существует полиномиальная зависимость между переменными x, ρ и y, ρ, т. е. полиномиальные уравнения

$$P(x, \rho) = 0, \quad Q(y, \rho) = 0. \tag{5.3}$$

Исключая из этих уравнений ρ, мы получим полиномиальное уравнение для кривой S_p

$$F(x, y) = 0. \tag{5.4}$$

Приведем несколько примеров уравнений (3) для различных значений p (напомним, что $x^2 + y^2 = \rho$):

$$p = 1, \qquad x = \frac{-1 + 4\rho^2 + \rho^4}{4\rho^2};$$

$$p = 2, \qquad x = \frac{1 - 12\rho^2 + 27\rho^4 + 4\rho^6}{12\sqrt{3}\,\rho^4};$$

$$p = 3, \qquad x = \frac{-1 + 24\rho^2 - 162\rho^4 + 256\rho^6 + 27\rho^8}{96\sqrt{3}\,\rho^6};$$

$$p = 1/2, \qquad x = \frac{-2 + 6\rho^2 + \rho^6}{3\sqrt{3}\,\rho^3};$$

$$p = 1/3, \qquad x = \frac{\sqrt{3}\,(-9 + 24\rho^2 - 2\rho^4 + 3\rho^8)}{32\,\rho^4}.$$

Внешний вид простейших кривых Серре показан на рис. 35.

Пусть $p = s/t$ — представление положительного рационального числа p в виде несократимой дроби ($s, t \in \mathbb{N}$). Можно показать, что соотношения (5.3) имеют вид

$$x = \frac{a_0 + a_1\rho^2 + \ldots + a_n\rho^{2n}}{b\rho^m}, \quad x^2 + y^2 = \rho^2, \tag{5.5}$$

где $n = s + t$ и $m = 2s + t - 1$. Кроме того, $a_0 = -t^3$, $a_n = s^s$. Поэтому для нечетных знаменателей t многочлен $F(x, y)$ из уравнения (5.4) выглядит следующим образом:

$$F(x,y) = bx\,(x^2 + y^2)^{s+(t-1)/2} - \\ - a_0 - a_1(x^2 + y^2) - \ldots - a_{s+t}\,(x^2 + y^2)^{s+t} \tag{5.6}$$

Степень кривой в этом случае равна $\alpha = 2n = 2s + 2t$. Для четных t многочлен $F(x, y)$ имеет более сложную структуру:

$$F(x,y) = b^2x^2\,(x^2 + y^2)^{2s+t-1} - \\ - [a_0 - a_1(x^2 + y^2) + \ldots + a_{s+t}(x^2 + y^2)^{s+t}]^2. \tag{5.7}$$

Его степень равна $4s + 4t$.

Мы будем рассматривать здесь лишь кривые Серре S_p с натуральными значениями параметра p (т. е. случай $t = 1$). Другие случаи пока не исследованы.

Хорошо известно, что лемниската является рациональной кривой, т. е. допускает рациональную параметризацию; иными словами, ее род равен нулю. Оказывается, это свойство выполняется и для всех кривых Серре S_p с натуральными значениями параметра p.

Теорема 3. *Для любого натурального числа p кривая Серре S_p рациональна.*

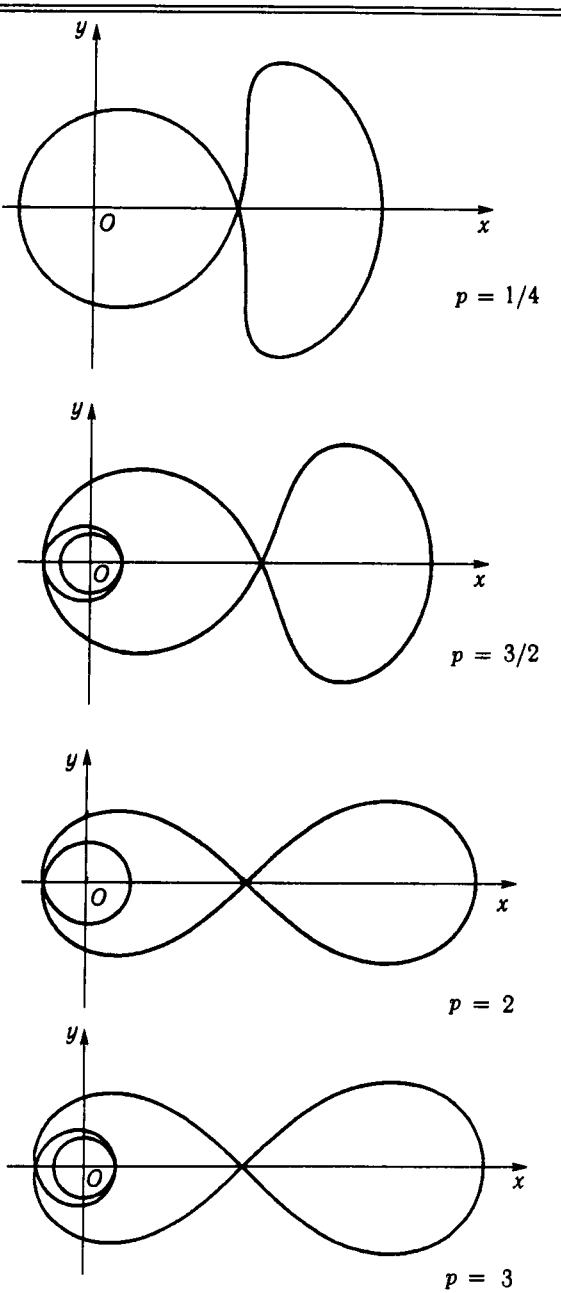


Рис. 35. Кривые Серре при различных параметрах p

Доказательство. Воспользуемся следующим утверждением (вытекающем, в свою очередь, из формулы Римана — Гурвица [Б14]): плоская алгебраическая кривая, задаваемая уравнением $F(x,y)=0$, является рациональной тогда и только тогда, когда обращается в нуль число

$$g=\frac{(d-1)(d-2)}{2}-\sum\frac{\nu(\nu-1)}{2},$$

называемое родом этой кривой. Здесь $d=\deg F$ и ν — кратность особой точки; суммирование ведется по всем особым точкам, включая и бесконечно удаленные. Итак, рассмотрим многочлен (5.6):

$$F(x,y)=bx(x^2+y^2)^p-a_0-a_1(x^2+y^2)-\ldots-a_{p+1}(x^2+y^2)^{p+1}.$$

Изучим, прежде всего, его поведение на бесконечности. С этой целью разложим его на однородные составляющие

$$F(x,y)=F_0(x,y)+\ldots+F_d(x,y),\qquad \alpha=2p+2, \tag{5.8}$$

и перейдем к однородному многочлену от переменных x, y, z

$$f(x,y,z)=z^dF_0(x,y)+\ldots+zF_{d-1}(x,y)+F_s(x,y). \tag{5.9}$$

Система уравнений для нахождения особых точек однородного многочлена $f(x,y,z)$ выглядит следующим образом:

$$\frac{\partial f}{\partial x}=\frac{\partial f}{\partial y}=\frac{\partial f}{\partial z}=f=0.$$

Следовательно, особые точки на бесконечно удаленной прямой $z=0$ имеют вид $(x:y:0)$, где (x,y) — решение системы уравнений

$$\frac{\partial F_d(x,y)}{\partial x}=\frac{\partial F_\alpha(x,y)}{\partial y}=F_{d-1}(x,y)=0. \tag{5.10}$$

Здесь

$$F_\alpha(x,y)=-a_{p+1}(x^2+y^2)^{p+1},\quad F_{d-1}(x,y)=bx(x^2+y^2)^p.$$

Таким образом, система уравнений (5.10) сводится к системе

$$x(x+2+y^2)^p=y(x^2+y^2)^p=0,$$

что дает ровно две точки $(\pm i:1:0)$.

Исследуем поведение кривой в этих особых точках. Для этого нужно рассмотреть однородный многочлен (5.9) в карте $y=1$ и перенести начало координат в точку $(x,z)=(\pm i,0)$ (подробности, относящиеся к используемой здесь технике, можно найти, например, в руководствах [Б29, Б14]. Поскольку вычисления в обоих случаях аналогичны, проведем их лишь для точки $(i:1:0)$, т. е. для $(x,z)=(i,0)$. В этом случае после указанных преобразований мы получаем многочлен

$$b(x+i)x^p(x+2i)^pz-a_0z^{2p+2}-a_1x(x+2i)z^{2p}-\ldots$$
$$\ldots-a_{p+1}x^{p+1}(x+2i)^{p+1}. \tag{5.12}$$

Начальная форма этого многочлена в точке $(0,0)$ равна

$$b'x^pz-a_0'z^{2p+2}n-a_1'xz^{2p}-\ldots-a_p'x^pz^2-a_{p+1}'x^{p+1}, \tag{5.13}$$

где коэффициенты b', $a_0',\ldots,a_{p+1}'$ вычисляются по коэффициентам b, $a_0,\ldots,a_{p+1}$. Таким образом, мы получаем особую точку кратности $p+1$ с характеристическими одночленами x^pz, z^{2p+2}, x^{p+1}. Отвечающая этой особой точке диаграмма Ньютона имеет вид, изображенный на рис. 36.

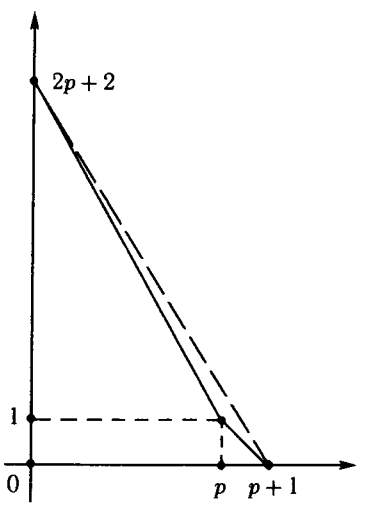


Рис. 36

Очевидно, что на сторонах этой диаграммы не содержится точек с целыми координатами. Это означает, что особенность состоит из двух ветвей. Чтобы вычислить понижение рода от этой особенности, ее можно разрешить раздутием. Для этого достаточно всего двух раздутий.

Очевидно, что на сторонах этой диаграммы не содержится точек с целыми координатами. Это означает, что особенность состоит из двух ветвей. Чтобы вычислить понижение рода от этой особенности, ее можно разрешить раздутием. Для этого достаточно всего двух раздутий.

Первое раздутие имеет вид $x\mapsto xz$, $z\mapsto x$. При этом

$$x^pz+z^{2p+2}+x^{p+1}\mapsto z^{p+1}(x^p+z^{p+1}+x^{p+1}).$$

В карте $x \mapsto x, \quad z \mapsto xz$ особенности нет, так как форма $x^p z + z^{2p+2} + x^{p+1}$ переходит при этом в $x^{p+1}(z + x^p z^{2p} + 1)$. В результате первого раздутия получаем одну особую точку типа $x^p + z^{p+1} + x^{p+1}$ кратности p.

При втором раздутии получаются простые точки. В самом деле, для $x \mapsto xz, \quad z \mapsto z$ имеем

$$x^p + z^{p+1} + x^{p+1} \mapsto z^p(x^p + z + x^{p+1}z).$$

Если же $x \mapsto x$, $z \mapsto xz$, то

$$x^p + z^{p+1} + x^{p+1} \mapsto x^p(1 + xz^{p+1} + x).$$

Дерево бесконечно близких точек имеет вид

$$(p+1) \longrightarrow (p) \longrightarrow (1),$$

поэтому вклад двух бесконечно удаленных точек $(\pm i : 1 : 0)$ в род g равен

$$-2\left[\frac{(p+1)p}{2} + \frac{p(p-1)}{2}\right] = -2p^2.$$

Следовательно, после разрешения этих особенностей род g кривой становится равным

$$g = \frac{(d-2)(d-2)}{2} - \sum{}' \frac{\nu(\nu-1)}{2} - 2p^2 = 2p - \sum{}' \frac{\nu(\nu-1)}{2} \leqslant p, \quad (5.14)$$

где $\sum'$ означает суммирование по всем особым точкам, расположенным в конечной части комплексной плоскости. В частности, из неравенства (5.14) вытекает, что вклад $\sum' \frac{\nu(\nu-1)}{2}$ от особых точек кривой S_p в конечной части плоскости не превосходит p.

Найдем теперь конечные особые точки кривой S_p. Основываясь на свойствах симметрии кривой, можно предположить, что эти особые точки расположены на оси Ox и, стало быть, удовлетворяют системе уравнений

$$\frac{\partial F(x,0)}{\partial x} = F(x,0) = 0. \quad (5.15)$$

Предположение это будет оправдано, если мы сможем убедиться, что вклад $\sum' \frac{\nu(\nu-1)}{2}$ от этих особых точек равен в точности p.

Чтобы проанализировать систему уравнений (5.15), удобно воспользоваться соотношением (5.1). Уравнение $F(x,0) = 0$ равносильно тому, что $\omega = 0$ или π, что, в свою очередь, равносильно уравнению

$$\cos(p\alpha - (p+1)\beta) = \pm 1, \quad (5.16)$$

откуда

$$p\alpha - (p+1)\beta = q\pi, \qquad q \in \mathbb{Z}. \quad (5.17)$$

Если добавить к (5.17) условия

$$\sin\beta = \sqrt{\frac{p}{p+1}}\sin\alpha \quad \text{и} \quad 0 \leqslant \alpha, \beta, \alpha+\beta \leqslant \pi, \quad (5.18)$$

мы получаем систему уравнений

$$\begin{cases} \rho = \dfrac{p}{p+1}\alpha - \dfrac{q}{p+1}\pi, \qquad q \in \mathbb{Z}, \\ \rho = \arcsin\left(\sqrt{\dfrac{p}{p+1}}\sin\alpha\right), \end{cases} \quad (5.19)$$

Для $q = 0$ прямая $\beta = \frac{p}{p+1}\alpha$ пересекает кривую $\beta = \arcsin\left(\sqrt{\frac{p}{p+1}}\sin\alpha\right)$ ровно в двух точках. Все остальные прямые $\beta = \frac{p}{p+1}\alpha\frac{q}{p+1}\pi$ либо вовсе не пересекают ее (при $q < 0$ и при $q > p$), либо пересекают ровно в одной точке (при $q = 1, \ldots, p$). Последняя точка пересечения имеет координаты $(\pi, 0)$. Таким образом, наша система (5.19) или, что то же самое, исходная система (5.15), имеет ровно две простых и p двойных корней. Это означает, что не существует других особых точек и что род кривой $g = 0$. Следовательно, все кривые S_p при натуральных p рациональны. □

Сопоставим кривой Серре S_p эллиптическую кривую. Для этого представим дифференциал дуги

$$dl = \sqrt{p}\,\frac{d\varphi}{\sqrt{1 - \frac{p}{p+1}\sin\varphi}}$$

кривой S_p в виде

$$dl = \frac{dx}{\sqrt{4x^3 - g_2 x - g_3}}.$$

Прямые вычисления показывают, что

$$g_2 = \frac{16p^2 + 16p + 1}{12p^2(p+1)^2}, \quad g_3 = -\frac{(2p+1)(32p^2+32p-1)}{6^3 p^3 (p+1)^3}.$$

Назовем кривую

$$y^2 = 4x^3 - g_2 x - g_3$$

эллиптической кривой, ассоциированной с кривой S_p.

Выше мы отмечали, что решающим моментом в доказательстве теоремы Абеля о делении лемнискаты является инвариантность решетки периодов лемнискатической функции $\varphi(\alpha)$ относительно умножения на комплексную единицу i. Эта инвариантность служит проявлением важного общего свойства некоторых эллиптических кривых — наличия у них так называемого комплексного умножения.

Пусть E — некоторая эллиптическая кривая. Как групповое многообразие она изоморфна фактору комплексной плоскости $\mathbb{C}$ по решетке Λ, натянутой на периоды ω_1, ω_2. Кривая E изоморфна фактору $\mathbb{C}$ по решетке $z\omega_1$, $z\omega_2$ для любого $z \in \mathbb{C}$, поэтому можно считать, что $\omega_1 = 1$, $\omega_2 = \tau$ и $\operatorname{Im}\tau > 0$. Любой эндоморфизм кривой E индуцирует умножение на такое комплексное число z, что $z\tau \in \Lambda$. Эндоморфизмы кривой E образуют кольцо $A(E)$, содержащее в качестве так называемых тривиальных эндоморфизмов кольцо целых чисел $\mathbb{Z}$. Остальные эндоморфизмы, если они существуют, задаются комплексными числами и называются *комплексным умножением*. Если $A(E) \neq \mathbb{Z}$, то говорят, что кривая E *имеет комплексное умножение*.

В общем случае кривая E не имеет комплексных умножений. Действительно, предположим, что число z определяет нетривиальный эндоморфизм E.

$$z = a + bi, \quad zi = c + di \qquad (a, b, c, d \in \mathbb{Z},\ b \neq 0)$$

и, следовательно,

$$ai + b\tau^2 = c + d\tau. \tag{5.20}$$

Таким образом, τ должно принадлежать мнимому квадратичному полю K, а z должно принадлежать кольцу целых элементов

$O(K)$ поля K, поскольку z определяет эндоморфизм $\mathbb{Z}$-модуля конечного ранга. Следовательно, $A(E)$ — это подкольцо в $O(K)$, содержащее $\mathbb{Z}$ и имеющее как $\mathbb{Z}$-модуль ранг два. Наоборот, любое такое подкольцо R в мнимом квадратичном поле K можно получить описанным способом. Для этого достаточно положить $E = \mathbb{C}/\mathbb{R}$.

Наличие у эллиптической кривой E комплексного умножения удобнее всего выражать с помощью так называемого j-инварианта. Если E задана уравнением в форме Вейерштрасса

$$y^2 = x^3 + ax + b,$$

то ее j-инвариант определяется формулой

$$j(E) = 1728\,\frac{4a^3}{4a^3 + 27b^2} \tag{5.21}$$

(подробнее о j-инварианте см. в гл. 7 § 17–19). Две эллиптические кривые изоморфны в том и только в том случае, когда их j-инварианты равны. Если кривая E определена над полем $\mathbb{Q}$, т. е. если $a, b \in \mathbb{Q}$, существует очень простой критерий наличия у нее комплексного умножения: кривая E обладает комплексным умножением тогда и только тогда, когда $j(E)$ — целое число [Б26, б; Б2]. Можно показать [Б2], что имеется ровно 13 классов эллиптических кривых с комплексным умножением и рациональным j-инвариантом. Значения j-инварианта для них таковы:

$$j = 2^6 \cdot 3^3,\quad 2^6 \cdot 5^3,\quad 0,\quad -3^3 \cdot 5^3,$$
$$-2^{15},\quad -2^{15} \cdot 3^3,\quad -2^{18} \cdot 3^3 \cdot 5^3,$$
$$-2^{15} \cdot 3^3 \cdot 5^3 \cdot 11^3,\quad -2^{18} \cdot 3^3 \cdot 5^3 \cdot 23^3 \cdot 29^3,\quad 2^3 \cdot 3^3 \cdot 11^3,$$
$$2^4 \cdot 3^3 \cdot 5^3,\quad 3^3 \cdot 5^3 \cdot 17^3,\quad -3 \cdot 2^{15} \cdot 5^3.$$

Теорема 4. *Среди эллиптических кривых, ассоциированных с кривыми Серре* S_p, $p \in \mathbb{N}$, *лишь эллиптическая кривая, соответствующая лемнискате* S_1, *обладает комплексным умножением.*

Доказательство. Из формулы длины дуги для кривых Серре

$$l = \sqrt{p} \int_0^{\varphi} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2 \varphi}}, \qquad k^2 = \frac{p}{p+1},$$

нетрудно получить, что $dl = \dfrac{dx}{y}$, где

$$y^2 = x^4 - 2(2p+1)x^2 + 1. \tag{5.22}$$

Это — уравнение эллиптической кривой в форме Лежандра. Переходя к уравнению в форме Вейерштрасса и вычисляя затем j-инвариант, получим

$$j = 2^8 \frac{(p^2+p+1)^3}{p^2(p+1)^2}. \tag{5.23}$$

Так как

$$j = 2^8 \frac{(p(p+1)+1)^3}{(p(p+1))^2} = 2^8 p(p+1) + 3\cdot 2^8 + \frac{3\cdot 2^8}{p(p+1)} + \frac{2^8}{p^2(p+1)^2},$$

то число j может быть целым лишь при $p = 1$. В этом случае $j = 2^6 \cdot 3^3$.

Таким образом, лишь эллиптическая кривая, отвечающая лемнискате S_1, допускает комплексное умножение. □

В заключение сформулируем несколько проблем. Прежде всего, было бы очень полезно найти для кривых Серре S_p с натуральным параметром p их рациональные параметризации, подобные параметризации лемнискаты. Однако главной проблемой является исследование структуры кривых Серре для нецелых значений параметра p. Здесь имеется два непохожих семейства: с нечетными и с четными значениями знаменателя t. Представляется очень правдоподобным, что при нечетных t все кривые Серре рациональны. В то же время компьютерные вычисления показывают, что при четных t могут появляться нерациональные кривые. Далее, было бы очень важно вычислить j-инварианты соответствующих эллиптических кривых и исследовать наличие у них комплексного умножения. И, наконец, было бы очень интересно проверить, не обобщается ли доказательство Эйзенштейна на кривые без комплексного умножения.

ГЛАВА 5

АРИФМЕТИКА ЭЛЛИПТИЧЕСКИХ КРИВЫХ

В этой главе мы рассмотрим некоторые наиболее известные диофантовы уравнения. *Диофантовым уравнением* называется полиномиальное уравнение

$$f(x_1, \ldots, x_n) = 0,$$

коэффициенты которого целые числа. Решение $(x_1, x_2, \ldots, x_n)$ назовем *целочисленным*, если все x_i — целые. Решение этого уравнения в рациональных числах называется *рациональным*. Очевидно, что в случае, когда f — однородный многочлен, задача нахождения рационального решения равносильна задаче нахождения целочисленного решения.

Истоки теории диофантовых уравнений восходят к античной математике. Еще Евклид разработал метод нахождения наибольшего общего делителя d двух чисел a_1 и a_2. Вообще говоря, этот метод приводит к решению диофантова уравнения $a_1x_1 + a_2x_2 = d$, однако сам Евклид решением таких уравнений не занимался. Евклид знал также способ нахождения наибольшего общего делителя трех чисел; этот способ годится и для n чисел. Он основан на том, что $((a_1, \ldots, a_{n-1}), a_n) = (a_1, \ldots, a_n)$.

При решении геометрических задач древнегреческие математики никогда не пользовались составлением уравнений, так как разработанное ими понятие геометрической величины не позволяло этого делать. Вследствие этого, у древнегреческих математиков не могло быть большого интереса к уравнениям. После Евклида новые приемы решения уравнений в целых числах были созданы лишь в III в. н. э. александрийским математиком Диофантом. Он разработал способы решения в целых и рациональных числах квадратных и некоторых кубических уравнений с двумя и более неизвестными, заложив тем самым фундамент новой математической дисциплины, называемой теперь в его честь теорией диофантовых уравнений. Свои исследования он изложил

в обширном сочинении в 13 книгах под названием «Арифметика». Удивительна судьба этого сочинения. После своего появления в свет оно исчезло более чем на тысячелетие и считалось безвозвратно утраченным. Лишь в 1464 г. немецкий ученый Региомонтан случайно обнаружил 6 из 13 книг «Арифметики». В первый раз она была напечатана в латинском переводе в 1575 г. После ее французского издания 1621 г., подготовленного Баше де Мезирьяком, она стала настольной книгой многих математиков, в числе которых были Рене Декарт, который первым стал систематически использовать уравнения для решения геометрических задач, и Пьер Ферма. Как раз на полях своего экземпляра «Арифметики» П. Ферма записал одно из самых знаменитых в истории математики замечаний: «Cubum autem in duos cubos, aut quadrato-quadratum in duos quadrato-quadratos, et generaliter nullam in infinitum ultra quadratum potestatem in duas ejusdem nominis fas est divedere; cujus rei demonstrationem mirabilem sane detexi. Hanc marginis exiguitas non caperet *)».

Совокупность приемов, разработанная Диофантом, сохранила свое значение до настоящего времени. Она носит название метода секущих. Этот метод позволяет полностью проанализировать уравнение второй степени и служит прототипом для исследования кубических уравнений.

§ 1. Метод секущих Диофанта. Диофантовы уравнения второй степени

Прежде чем рассмотреть общую ситуацию, продемонстрируем метод секущих на конкретном примере — одном из тех, которые Диофант разбирает в своей «Арифметике». Пусть дано уравнение

$$x^2 - y^2 = 1 \tag{1.1}$$

и требуется найти все его рациональные решения. Уравнение (1.1) задает на плоскости Oxy гиперболу. Сразу видно, что

*) «Невозможно разложить куб на два куба, или биквадрат на два биквадрата, или вообще степень, большую двух, на две степени с тем же самым показателем; я нашел этому поистине чудесное доказательство, однако поля слишком малы, чтобы оно здесь уместилось». (*Перев. с лат.*)

точка $P = (1,0)$ — точка пересечения гиперболы с осью Ox — является решением. Проведем через эту точку секущую (рис. 37)

$$y = k(x - 1) \tag{1.2}$$

и найдем вторую точку пересечения этой секущей с кривой (1.1). Для этого подставим выражение (1.2) в уравнение (1.1) и решим полученное квадратное уравнение относительно x. В результате получим

$$x_{1,2} = \frac{-k^2 \pm 1}{1 - k^2}.$$

Один корень $x_1 = 1$ нам известен — он соответствует точке $(1,0)$, другой корень x_2 определяет вторую точку

$$\left(\frac{k^2+1}{k^2-1}, \frac{2k}{k^2-1}\right). \tag{1.3}$$

Для любого рационального $k \neq \pm 1$ эта формула определяет точку на нашей кривой, а значит, и рациональное решение данного уравнения. При $k = \pm 1$ секущая пересекает кривую только в точке P (рис. 37). Наоборот, для любого рационального решения, т. е. рациональной точки M на кривой, секущая PM задается уравнением (1.2) с рациональным k, так как тогда катеты прямоугольного треугольника PMH рациональны. Таким образом, формула (1.3) при всевозможных рациональных $k \neq \pm 1$ дает все решения уравнения (1.1) в рациональных числах.

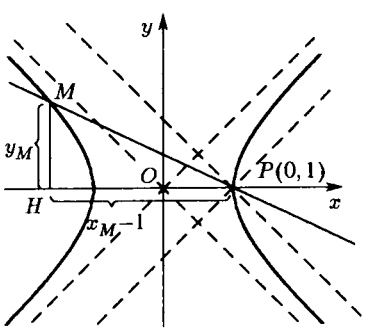


Рис. 37

Этот метод применим не только к многочлену $x^2 - y^2 - 1$, но и к любому многочлену второй степени от двух переменных

$$p(x,y) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F$$

с целыми или рациональными коэффициентами при условии, что на кривой $p(x,y) = 0$ есть хотя бы одна рациональная точка $P_0 = (x_0, y_0)$. В самом деле, проведем через эту точку прямую

$$y - y_0 = k(x - x_0) \tag{1.4}$$

и найдем точки пересечения кривой $p(x,y)=0$ с этой прямой. Для этого подставим (1.4) в уравнение $p(x,y)=0$. Многочлен

$$p(x, y_0 + k(x - x_0))$$

имеет степень 2 относительно x. Если положить

$$p(x, y_0 + k(x - x_0)) = r(k)x^2 + s(k)x + t(k),$$

где $r(k), s(k), t(k)$ — многочлены от k, то для x получим уравнение

$$r(k)x^2 + s(k)x + t(k) = 0.$$

Нам известен один из корней этого уравнения, а именно $x = x_0$. Поэтому другой корень $x = x_1$ можно найти с помощью соотношения

$$x_0 + x_1 = -\frac{s(k)}{r(k)}.$$

Подставив выражение

$$x_1 = -x_0 - \frac{s(k)}{r(k)}$$

в уравнение (1.4), получим

$$y_1 = y_0 + k\left(-2x_0 + \frac{s(k)}{r(k)}\right).$$

Таким образом, координаты второй точки имеют вид

$$(x_1, y_1) = \left(-x_0 - \frac{s(k)}{r(k)}, y_0 - k\left(2x_0 + \frac{s(k)}{r(k)}\right)\right). \qquad (1.5)$$

Так как $s(k)$ и $r(k)$ — многочлены с рациональными коэффициентами, то при любых рациональных k, для которых $r(k) \neq 0$, формула (1.5) дает рациональное решение уравнения $p(x,y)=0$, а поскольку при рациональных x_1 и y_1 угловой коэффициент

$$k = \frac{y_1 - y_0}{x_1 - x_0}$$

также рационален, формула (1.5) дает все рациональные решения при условии, что заранее известно хотя бы одно рациональное решение.

Вопрос о существовании рациональной точки на кривой второго порядка $p(x,y)=0$ довольно сложен. Такие точки имеются

далеко не всегда. Например, их нет на окружности $x^2 + y^2 = 3$ и на эллипсе $x^2 + 82y^2 = 3$. Если кривая $p(x,y)=0$ приводима над $\mathbb{Q}$, то задача нахождения рациональной точки сводится к исследованию линейного диофантова уравнения, поэтому можно ограничиться рассмотрением лишь неприводимых над $\mathbb{Q}$ кривых

$$p(x,y) = Ax^2 + 2Bxy + Cy^2 + Dx + Ey + F = 0.$$

Хорошо известно, что в случае, когда дискриминант $\Delta = AC - B^2$ квадратичной формы $Ax^2 + 2Bxy + Cy^2$ равен нулю, уравнение $p(x,y)=0$ с помощью невырожденного линейного преобразования с рациональными коэффициентами приводится либо к виду $ax^2 + y = 0$, где $a \neq 0$, либо к виду $x^2 - c = 0$, где c не является полным квадратом. В случае же, когда $\Delta \neq 0$, уравнение $p(x,y)=0$ приводится к виду

$$ax^2 + by^2 + c = 0, \quad ab \neq 0. \qquad (1.6)$$

Очевидно, что кривая $x^2 - c = 0$ не имеет рациональных точек. Очевидно также, что на кривой $ax^2 + y = 0$ есть рациональная точка $(0,0)$. Стало быть, остается исследовать кривую (1.6). Если $c = 0$, то $(0,0)$ — рациональная точка кривой (1.6), поэтому можно считать, что $abc \neq 0$. Прежде всего, для разрешимости уравнения (1.6) в рациональных числах необходимо, чтобы не все коэффициенты a, b и c были одного знака. Выполнив при необходимости замену переменных

$$x \mapsto x^{-1},\ y \mapsto yx^{-1} \quad \text{или} \quad x \mapsto xy^{-1},\ y \mapsto y^{-1},$$

уравнение (1.6) можно привести к виду

$$ax^2 + by^2 - c = 0, \qquad a > 0,\ b > 0,\ c > 0. \qquad (1.7)$$

При этом можно считать, что a, b и c — взаимно простые в совокупности и свободные от квадратов целые числа.

Если $x = p/r$, $y = q/r$, где $p, q, r \in \mathbb{Z}$ — некоторое рациональное решение уравнения (1.7), то уравнение

$$ax^2 + by^2 - cz^2 = 0 \qquad (1.8)$$

обладает ненулевым целочисленным решением $(x,y,z) = (p,q,r)$. Обратно, если уравнение (1.8) имеет нетривиальное

решение в целых числах (x, y, z), то $z \neq 0$ (иначе мы получили бы нетривиальное решение уравнения $ax^2 + by^2 = 0$, $a > 0$, $b > 0$) и тогда уравнение (1.7) разрешимо в рациональных числах. Следовательно, вопрос о существовании рациональных решений уравнения (1.6) сводится при выполнении указанного выше необходимого условия к вопросу о нетривиальной разрешимости в целых числах уравнения (1.8). При этом числа a, b и c можно считать не только взаимно простыми в совокупности, но и попарно взаимно простыми. В самом деле, пусть, например, a и b делятся на q. Тогда

$$q(ax^2 + by^2 - cz^2) = a_1x_1^2 + b_1y_1^2 - c_1z^2,$$

где $x_1 = qx$, $y_1 = qy$, $a_1 = a/q$, $b_1 = b/q$, $c_1 = qc$.

Критерий существования нетривиального решения уравнения (1.8) был получен еще Лежандром.

Теорема 1. *Если a, b и c — попарно взаимно простые натуральные числа, свободные от квадратов, то уравнение*

$$ax^2 + by^2 - cz^2 = 0$$

имеет нетривиальное целочисленное решение тогда и только тогда, когда разрешимы все три сравнения:

$$x^2 - bc \equiv 0 \pmod a,$$
$$x^2 - ac \equiv 0 \pmod b,$$
$$x^2 + ab \equiv 0 \pmod c.$$

Нам будет удобнее перейти от уравнения (1.8) к уравнению

$$acx^2 + bcy^2 = z_1^2.$$

Если (x, y, z_1) — решение этого уравнения, то z_1 делится на c, так как c свободно от квадратов. Поэтому $(x, y, z_1/c)$ — решение уравнения (1.8). Таким образом, вместо теоремы 1 достаточно доказать следующее утверждение.

Теорема 2. *Пусть a и b — натуральные числа, свободные от квадратов. Уравнение*

$$ax^2 + by^2 = z^2 \tag{1.9}$$

имеет нетривиальное решение в целых числах тогда и только тогда, когда существуют такие целые числа α, β и γ, что

а) $\alpha^2 - a \equiv 0 \pmod b$,

б) $\beta^2 - b \equiv 0 \pmod a$,

в) $\gamma^2 + ab/h^2 \equiv 0 \pmod h$,

где $h = (a, b)$.

Доказательство. Предположим сначала, что уравнение (1.9) имеет нетривиальное целочисленное решение x, y, z. Можно считать что $(x, y, z) = 1$. Из (1.9) следует, что

$$ax^2 \equiv z^2 \pmod b.$$

Пусть $(b, x) = d$. Тогда $x = dx_1$ и $b = db_1$, поэтому

$$ad^2x_1^2 + db_1y^2 = z^2.$$

Следовательно, $z = dz_1$. Значит, b_1y^2 делится на d, причем d свободно от квадратов, так как b свободно от квадратов. В итоге получаем, что все числа x, y, z делятся на d, т. е. $d = 1$. Поэтому существует такое число x', что

$$xx' \equiv 1 \pmod b.$$

Тогда

$$a \equiv \alpha^2 \pmod b,$$

где $\alpha = x'z$. Аналогично доказывается разрешимость сравнения

$$b \equiv \beta^2 \pmod a.$$

Сравнение в) можно записать в виде

$$\gamma^2 + a_1b_1 \equiv 0 \pmod h,$$

где $a = ha_1, b = hb_1$. При этом числа a_1, b_1, h попарно взаимно простые и свободные от квадратов. Так как a и b делятся на h, то $z = hz_1$, а значит, $a_1x^2 + b_1y^2 = hz_1^2$. Если $(x, h) = d$, то y и z_1 делятся на d. Поэтому $d = 1$ и существует такое число x', что

$$xx' \equiv 1 \pmod h.$$

Домножив сравнение

$$a_1x^2 + b_1y^2 \equiv 0 \pmod{h}$$

на $b_1x'^2$, получим

$$a_1b_1 + \gamma^2 \equiv 0 \pmod{h},$$

где $\gamma = b_1x'y$.

Предположим теперь, что разрешимы все три сравнения а)–в). Если $a = 1$, то теорема очевидна. Кроме того, можно считать, что $a \geqslant b$, поскольку при $b > a$ следует лишь поменять местами x и y. Пусть $a = b$, тогда согласно в)

$$\gamma^2 + 1 \equiv 0 \pmod{a}.$$

Предположим, что один из простых делителей числа a имеет вид $p = 4k + 3$. Тогда

$$\gamma^2 + 1 \equiv 0 \pmod{p}$$

и

$$(\gamma^2)^{\frac{p-1}{2}} \equiv (-1)^{\frac{p-1}{2}} \equiv (-1)^{2k+1} \equiv -1 \pmod{p}.$$

С другой стороны, согласно малой теореме Ферма,

$$(\gamma^2)^{\frac{p-1}{2}} \equiv \gamma p - 1 \equiv 1 \pmod{p}.$$

Приходим к противоречию, поэтому a является произведением простых чисел вида $4k+1$ и, возможно, числа 2. Каждый простой сомножитель числа a представим в виде суммы двух квадратов (лемма 4 § 4, гл. 4). Ясно также, что

$$(x^2 + y^2)(z^2 + t^2) = (xz - yt)^2 + (xt + yz)^2.$$

Поэтому $a = r^2 + s^2$ для некоторых целых r и s. Тогда решение уравнения (1.9) задается в виде

$$x = r, \quad y = s, \quad z = r^2 + s^2.$$

Остается рассмотреть случай, когда $a > b > 1$. От уравнения (1.9) перейдем к новому уравнению $Ax^2 + by^2 = z^2$, где $0 < A < a$ и для A разрешимы сравнения, аналогичные а)–в). После конечного числа шагов, меняя местами A и b при $A < b$, мы получим либо $A = b$, либо $b = 1$. В каждом из этих случаев,

как уже было показано, решение существует. По этому решению мы построим решение уравнения (1.9).

Согласно б) существует такое число β, что $\beta^2 - b = aAk^2$, где число A свободно от квадратов. При этом можно считать, что $|\beta| \leqslant a/2$. Покажем сначала, что $0 < A < a$. Так как

$$\beta^2 = aAk^2 + b < a(Ak^2 + 1),$$

то $A \geqslant 0$. Кроме того, $b \neq 1$ и b свободно от квадратов. Поэтому $\beta^2 \neq b$, а значит, $A \neq 0$. Ясно также, что

$$A = \frac{1}{ak^2}(\beta^2 - b) < \frac{\beta^2}{ak^2} \leqslant \frac{\beta^2}{a} \leqslant \frac{a}{4} < a.$$

Докажем теперь разрешимость требуемых сравнений. Так как $\beta^2 - b = aAk^2$, то

$$\beta^2 \equiv b \pmod{A},$$

поэтому разрешимо сравнение, аналогичное б). Кроме того, $\beta = h\beta_1$, где $h = (a, b)$, поэтому

$$h\beta_1^2 - b_1 = a_1Ak^2. \tag{1.10}$$

В частности,

$$h\beta_1^2 \equiv a_1Ak^2 \pmod{b_1}.$$

Из а) следует, что

$$a_1 \equiv h\alpha_1^2 \pmod{b_1}.$$

Поэтому

$$h\beta_1^2 \equiv hA(\alpha_1k)^2 \pmod{b_1}.$$

А так как h, k и a_1 взаимно просты с b_1, то

$$A \equiv p^2 \pmod{b_1}.$$

Кроме того, из (1.10) следует, что

$$a_1Ak^2 \equiv -b_1 \pmod{h}.$$

Поэтому

$$A(a_1k)^2 \equiv -a_1b_1 \equiv \gamma^2 \pmod{h}.$$

Значит,

$$A \equiv q^2 \pmod{h}.$$

Числа h и b_1 взаимно просты, поэтому $hu+b_1v=1$ для некоторых целых чисел u и v. Рассмотрим число $x=hup+b_1vq$. Так как

$$hu\equiv 1 \pmod{b_1} \quad \text{и} \quad b_1v\equiv 1 \pmod{h},$$

то

$$x\equiv p \pmod{b_1} \quad \text{и} \quad x\equiv q \pmod{h},$$

а значит, $A\equiv x^2 \pmod{b_1}$ и $A\equiv x^2 \pmod{h}$. Следовательно,

$$A\equiv x^2 \pmod{b}.$$

Рассмотрим теперь $H=(A,b)$. Пусть $A=HA_2$ и $b=Hb_2$. Тогда $\beta=H\beta_2$ и $H\beta_2^2-b_2=aA_2k^2$. Поэтому

$$-A_2b_2\equiv a(A_2k)^2 \pmod{H}.$$

Из в) следует, что

$$a\equiv \alpha^2 \pmod{H}.$$

Значит, $-A_2b_2\equiv r^2 \pmod{H}$.

Предположим теперь, что уравнение $AX^2+bY^2=Z^2$ имеет нетривиальное целочисленное решение. Умножим равенство $AX^2=Z^2-bY^2$ на $aAk^2=\beta^2-b$. В результате получим

$$a(AXk)^2=(Z^2-bY^2)(\beta^2-b)=(\beta Z+b)^2-b(\beta Y+Z)^2.$$

Значит, уравнение (1.9) имеет целочисленное решение $x=AXk$, $y=\beta Y+Z$, $z=\beta Z+bY$. Это завершает доказательство, так как $X\neq 0$. □

В 1950 г. канадский математик Хольцер [B21] получил уточнение теоремы Лежандра, которое дает эффективный алгоритм для нахождения нетривиального целочисленного решения уравнения (1.8), а значит, и для нахождения рационального решения уравнения (1.7).

Теорема 3 (Хольцер). *Пусть a, b, c, — попарно взаимно простые числа, свободные от квадратов, причем $a, b>0$, $c<0$. Тогда если уравнение $ax^2+by^2+cz^2=0$ имеет ненулевое целочисленное решение, то оно имеет также ненулевое целочисленное решение, для которого*

$$x^2\leqslant b|c|, \quad y^2\leqslant a|c|, \quad z^2\leqslant ab.$$

Доказательство (Морделл [B13, в]). Достаточно доказать, что $z^2\leqslant ab$. В самом деле, тогда $ax^2+by^2\leqslant ab|c|$, поэтому $x^2\leqslant b|c|$ и $y^2\leqslant a|c|$. Рассмотрим решение (x_0, y_0, z_0), для которого $(x_0, y_0)=1$ и $z_0^2>ab$. Достаточно показать, что с помощью этого решения можно построить новое решение (x, y, z), для которого $|z|<|z_0|$.

Тройка чисел

$$x_0+tx_1, \quad y_0+ty_1, \quad z_0+tz_1$$

будет решением тогда и только тогда, когда числа t, x_1, y_1, z_1 удовлетворяют соотношению

$$(ax_1^2+by_1^2+cz_1^2)t^2+2(ax_0x_1+by_0y_1+cz_0z_1)t=0.$$

При $t\neq 0$ на t можно сократить. В результате получим $t=m/n$, где

$$m=-2(ax_0x_1+by_0y_1+cz_0z_1), \quad n=ax_1^2+by_1^2+cz_1^2.$$

После умножения на знаменатель n получим решение

$$x_0n+x_1m, \quad y_0n+y_1m, \quad z_0n+z_1m.$$

Покажем,что все эти три числа делятся на

$$\Delta=(c,\ x_1y_0-y_1x_0).$$

По условию $(c, ab)=1$ и $(x_0, y_0)=1$. Поэтому из равенства

$$ax_0^2+by_0^2+cz_0^2=0$$

следует, что $(c, abx_0y_0)=1$, а значит, $(\Delta, abx_0y_0)=1$. Следовательно, существуют такие целые числа x_0^{-1} и y_0^{-1}, что

$$x_0x_0^{-1}\equiv 1 \pmod{\Delta}$$

и

$$y_0y_0^{-1}\equiv 1 \pmod{\Delta}.$$

А так как

$$x_1y_0-y_1x_0\equiv 0 \pmod{\Delta},$$

то

$$x_1\equiv y_1x_0y_0^{-1} \pmod{\Delta}.$$

Теперь легко показать, что

$$P = ax_0x_1 + by_0y_1 \equiv 0 \pmod{\Delta},$$
$$Q = ax_1^2 + by_1^2 \equiv 0 \pmod{\Delta}.$$

В самом деле,

$$P \equiv ax_0(y_1x_0y_0^{-1}) + by_0y_1 \equiv y_1(ax_0^2 + by_0^2)\,y_0^{-1} \equiv 0 \pmod{\Delta},$$
$$Q \equiv a(y_1x_0y_0^{-1})^2 + by_1 \equiv (ax_0^2 + by_0^2)y_1^2y_0^{-2} \equiv 0 \pmod{\Delta}.$$

Таким образом,

$$x_0n + x_1m = x_0(ax_1^2 + by_1^2 + cz_1^2) - 2x_1(ax_0x_1 + by_0y_1 + cz_0z_1) \equiv$$
$$\equiv x_0Q + cx_0z_1^2 - 2x_1P - 2cx_1z_0z_1 \equiv 0 \pmod{\Delta}.$$

Аналогично доказывается, что и остальные числа делятся на Δ.

Пусть δ — делитель числа Δ. Тогда тройка чисел

$$x = (x_0n + x_1m)\,\delta^{-1},$$
$$y = (y_0n + y_1m)\,\delta^{-1},$$
$$z = (z_0n + z_1m)\,\delta^{-1}$$

является целочисленным решением. Его нужно подобрать так, чтобы выполнялось неравенство $|z| < |z_0|$. Ясно, что

$$-\frac{\delta z}{cz_0} = \left(z_1 + \frac{ax_0x_1 + by_0y_1}{cz_0}\right)^2 - \left(\frac{ax_0x_1 + by_0y_1}{cz_0}\right)^2 - \frac{ax_1^2 + by_1^2}{c}.$$

Учитывая, что $cz_0^2 = -(ax_0^2 + by_0^2)$, получаем

$$-\frac{\delta z}{cz_0} = \left(z_1 + \frac{ax_0x_1 + by_0y_1}{cz_0}\right)^2 + \frac{ab}{c^2z_0^2}(y_0x_1 - x_0y_1)^2. \qquad (1.11)$$

Теперь построим требуемое решение. В качестве x_1 и y_1 возьмем произвольное решение уравнения $y_0x_1 - x_0y_1 = \delta$. Тогда на δ накладывается лишь одно условие — δ делит c.

Первый случай: число c четно. В качестве δ возьмем число $c/2$, а z_1 выберем так, чтобы выполнялось неравенство

$$\left|z_1 + \frac{ax_0x_1 + by_0y_1}{cz_0}\right| \leqslant \frac{1}{2}.$$

Тогда из (1.11) получаем

$$\frac{1}{2}\left|\frac{z}{z_0}\right| \leqslant \frac{1}{4} + \frac{ab}{4z_0^2}.$$

По условию $z_0^2 > ab$, поэтому $|z| < |z_0|$.

Второй случай: число c нечетно. Потребуем, чтобы выполнялось условие

$$zx_1 + by_1 + cz_1 \equiv 0 \pmod 2; \qquad (1.12)$$

Иными словами, число z_1 должно иметь ту же четность, что и число $ax_1 + by_1$.

Соотношение (1.12) эквивалентно тому, что

$$n = ax_1^2 + by_1^2 + cz_1^2 = 0 \pmod 2.$$

Следовательно, числа $x_0n + x_1m$, $y_0n + y_1m$, $z_0n + z_1m$ делятся на 2δ. Поэтому если вместо δ взять число $\delta' = 2\delta$, то для него тоже будет выполняться соотношение (1.11). В этом случае оно принимает вид

$$-\frac{2\delta z}{cz_0} = \left(z_1 + \frac{ax_0x_1 + by_0y_1}{cz_0}\right)^2 + \frac{ab}{c^2z_0^2}(y_0x_1 - x_0y_1)^2.$$

Положим $\delta = c$, а число z_1 выберем так, чтобы выполнялось соотношение (1.12) и при этом имело место неравенство

$$\left|z_1 + \frac{ax_0x_1 + by_0y_1}{cz_0}\right| \leqslant 1.$$

Тогда

$$2\left|\frac{z}{z_0}\right| \leqslant 1 + \frac{ab}{z_0^2} < 2,$$

т. е. $|z| < |z_0|$. □

Задачи

1. Какие из следующих уравнений имеют нетривиальные целочисленные решения?

а) $3x^2 - 5y^2 + 7z^2 = 0$.
б) $7x^2 + 11y^2 - 19z^2 = 0$.
в) $8x^2 - 5y^2 - 3z^2 = 0$.
г) $11x^2 - 3y^2 - 41z^2 = 0$.

2. Найдите рациональные точки следующих кривых:
а) $x^2 - 3y^2 = 1$;
б) $x^2 + 2y^2 = 9$;
в) $x^2 - 6y^2 = 1$.

3. Докажите, что уравнение $x^2 - 2y^2 = 3$ не имеет целочисленных решений.

4. Пусть d — натуральное число, свободное от квадратов.
а) Докажите, что если (x_1, y_1) — целочисленное решение уравнения $x^2 - dy^2 = 1$ и $(x_1 + y_1\sqrt{d})^n = x_n + y_n\sqrt{d}$, где x_n и y_n — целые числа, то (x_n, y_n) тоже решение этого уравнения.
б) Докажите, что если уравнение $x^2 - dy^2 = -1$ имеет целочисленное решение, то уравнение $x^2 - dy^2 = 1$ тоже имеет целочисленное решение.
в) Докажите, что уравнение $x^2 - dy^2 = 1$ имеет хотя бы одно целочисленное решение.
г) Докажите, что если уравнение $x^2 - dy^2 = n$, где $n \neq 0$, имеет хотя бы одно целочисленное решение, то оно имеет их бесконечно много.

§ 2. Сложение точек на кубической кривой

В своей «Арифметике» Диофант не ограничился уравнениями второй степени. Он с успехом решает и некоторые кубические уравнения, демонстрируя общий прием при нахождении рациональных решений уравнения

$$y(6 - y) = x^3 - x.$$

Однако систематический интерес к диофантовым уравнениям третьей степени возник, по-видимому, лишь в связи с античной задачей о конгруэнтных числах, исследование которой было предпринято арабскими математиками X в.

Положительное число $r \in \mathbb{Q}$ называется *конгруэнтным*, если оно является площадью некоторого прямоугольного треугольника с рациональными длинами сторон. Например, 6 — площадь треугольника со сторонами 3, 4 и 5; стало быть, 6 — конгруэнтное число. Пусть число r конгруэнтно и $a, b, c \in \mathbb{Q}$ — стороны прямоугольного треугольника с площадью r. Для любого $r \in \mathbb{Q}$ можно найти такое $s \in \mathbb{Q}$, что $s^2 r$ — целое число, свободное от квадратов. Но площадь треугольника со сторонами sa, sb и sc равна $s^2 r$. Таким образом, без ограничения общности можно считать, что $r = n$ — натуральное число, свободное от квадратов. Следует отметить, что в определении конгруэнтного числа стороны треугольника должны быть лишь рациональными числами, а не обязательно натуральными. В то время как 6 — наименьшая возможная площадь треугольника с целочисленными сторонами, существует прямоугольный треугольник с рациональными сторонами и с площадью 5. Такую площадь имеет треугольник со сторонами $(3/2, 20/3, 41/6)$. Можно показать, что 5 — наименьшее конгруэнтное натуральное число.

Оказывается, что проблема описания всех конгруэнтных чисел сводится к диофантову уравнению третьей степени. А именно, справедливо следующее утверждение.

Теорема 1. *Пусть n — натуральное число, свободное от квадратов. Тогда эквивалентны следующие три условия:*

1) *n — конгруэнтное число;*

2) *существует такое рациональное число x, что числа $x, x + n$ и $x - n$ являются квадратами рациональных чисел;*

3) *на кривой $y^2 = x^3 - n^2 x$ существует рациональная точка (x, y), координата x которой является квадратом рационального числа, причем знаменатель x — четное число, а числитель x не имеет общих делителей с n.*

Доказательство. Сначала докажем эквивалентность условий (1) и (2). Пусть $a < b < c$ — такая тройка положительных рациональных чисел, что $a^2 + b^2 = c^2$ и $n = ab/2$. Положим $x = c^2/4$. Тогда $x + n = (a+b)^2/4$ и $x - n = (a-b)^2/4$. Следовательно, $x, x + n$ и $x - n$ — квадраты рациональных чисел. Наоборот, если x удовлетворяет условию (2), то положим $c = 2\sqrt{x}$ и найдем числа a и b из системы уравнений

$$\begin{cases} (a+b)^2 = 4(x+n), \\ (a-b)^2 = 4(x-n). \end{cases}$$

Иными словами, если x удовлетворяет условию (2), то искомый треугольник имеет стороны

$$a = \sqrt{x+n} - \sqrt{x-n}, \quad b = \sqrt{x+n} + \sqrt{x-n}, \quad c = 2\sqrt{x}.$$

Чтобы доказать эквивалентность условий (2) и (3), рассмотрим такое рациональное число x, что $x, x+n$ и $x-n$ — квадраты

рациональных чисел. Тогда $x = u^2$ и $(x+n)(x-n) = v^2$. Положим $y = uv$. Тогда

$$y^2 = x^3 - n^2x, \tag{2.1}$$

т. е. точка (x, y) лежит на кубической кривой (2.1). Поскольку $x = c^2/4$, знаменатель x делится на 2. А так как n свободно от квадратов, то числитель x не имеет общих делителей с n.

Наоборот, если $x = u^2 = (c/2)^2$ и $x^3 - n^2x = y^2$, то

$$v^2 = y^2/x = x^2 - n^2 = (x+n)(x-n)$$

и мы имеем пифагорову тройку $v^2 + n^2 = x^2$. Числа x^2 и v^2 имеют один и тот же знаменатель q^4, причем число q по предположению четно. Следовательно, числа q^2v и q^2x целые, а число q^2n четное, причем q^2x и q^2n не имеют общих делителей и

$$(q^2v)^2 + (q^2n)^2 = (q^2x)^2.$$

Поэтому $q^2v = s^2 - t^2$, $q^2n = 2st$, $q^2x = s^2 + t^2$, где s и t — целые числа. Так как

$$\left(\frac{2s}{q}\right)^2 + \left(\frac{2t}{q}\right)^2 = 4x = (2u)^2,$$

то треугольник со сторонами $2s/q$, $2t/q$ и $2u$ прямоугольный, причем его площадь равна $2st/q^2 = q^2n/q^2 = n$. □

Задача о конгруэнтных числах, а также некоторые другие классические задачи, например, задача об отыскании рациональных решений уравнения $x^3 + y^3 = 1$, являются частными случаями проблемы нахождения рациональных решений общего кубического уравнения $f(x, y) = 0$ с двумя неизвестными, т. е. нахождения рациональных точек кривой C, заданной уравнением $f(x, y) = 0$.

Предположим сначала, что кубическая кривая $f(x, y) = 0$ имеет особую точку O, причем эта точка рациональна. В особой точке прямые пересекают кривую не менее чем двукратно. Из этого, в частности, следует, что кубическая кривая не может иметь двух особых точек O и O_1, так как иначе прямая OO_1 имела бы с кубической кривой не менее чем четырехкратное пересечение.

Проведем через точку $O = (x_0, y_0)$ рациональную прямую $x = x_0 + at$, $y = y_0 + bt$, где $a, b \in \mathbb{Q}$. Точки пересечения этой прямой с кубической кривой соответствуют корням многочлена

$F(t) = f(x_0 + at, y_0 + bt)$. Коэффициенты многочлена F рациональны, причем для почти всех прямых его степень равна трем (степень меньше трех лишь для прямых, проходящих через бесконечно удаленные точки кривой). Многочлен F имеет двукратный корень $t = 0$, соответствующий точке O. Следовательно, третий корень многочлена F рационален, т. е. он соответствует рациональной точке кривой. Ясно также, что прямая, соединяющая точку O с рациональной точкой кривой, задается уравнением с рациональными коэффициентами. Это дает полное описание множества рациональных точек особой кубической кривой.

В дальнейшем мы будем рассматривать лишь неособые кубические кривые. Напомним, как в главе 1 было определено сложение точек неособой кубической кривой. Пусть E — фиксированная точка данной кривой, A и B — некоторые точки кривой, X — точка пересечения прямой AB с кривой. Суммой точек A и B называется точка пересечения прямой EX с данной кривой. Легко проверить, что если кубическая кривая задается многочленом с целыми коэффициентами и точка E рациональна, то сумма двух рациональных точек будет рациональной точкой.

Кривая $y^2 = x^3 + ax + b$ будет неособой тогда и только тогда, когда ее дискриминант $\Delta = -(4a^3 + 27b^2)$ отличен от нуля. В том случае, когда в качестве нулевого элемента выбрана бесконечно удаленная точка, для такой кривой легко получить явные формулы сложения точек. Пусть прямая $y = px + q$ пересекает данную кривую в точках (x_i, y_i), $i = 1, 2, 3$. Тогда

$$(px + q)^2 = x^3 + ax + b$$

при $x = x_1, x_2, x_3$. Сумма корней этого уравнения равна p^2, следовательно,

$$x_3 = -x_1 - x_2 + p^2 = -x_1 - x_2 + \left(\frac{y_1 - y_2}{x_1 - x_2}\right)^2,$$

$$y_3 = px + q = \frac{y_1 - y_2}{x_1 - x_2}(x_3 - x_1) + y_1.$$

Ясно, что сумма точек (x_1, y_1) и (x_2, y_2) имеет координаты $(x_3, -y_3)$. В случае, когда $x_1 = x_2$, $y_1 = y_2$, достаточно заметить, что

$$\lim_{x_2 \to x_1} \frac{y_1 - y_2}{x_1 - x_2} = y'(x_1) = \frac{3x_1^2 + a}{2y_1}.$$

Отдельно следует разобрать случай $x_1 = x_2$, $y_1 \neq y_2$. В этом случае сумма точек есть бесконечно удаленная точка кривой.

Полученные формулы показывают, что, зная одну рациональную точку P кривой $y^2 = x^3 + ax + b$, можно найти рациональные точки $2P$, $3P$ и т. д. Рассмотрим, например, кривую $y^2 = x^3 - 2$ и точку $P = (3,5)$. Тогда

$$2P = \left(\frac{129}{100}, -\frac{383}{1000}\right)$$

— новая рациональная точка. Теперь можно вычислить $3P$, $4P$ и т. д. При этом с каждым шагом объем вычислений стремительно возрастает. Если обозначить через x_n первую координату точки nP, то

$$x_1 = 3, \quad x_2 = \frac{129}{100}, \quad x_3 = \frac{164323}{29241},$$

$$x_4 = \frac{2340922881}{58675600}, \quad x_5 = \frac{307326105747363}{160280942564521}.$$

А числитель x_{11} имеет уже 71 знак.

Следует отметить, что точки P, $2P$, $3P, \ldots$ не обязательно различны. В таком случае наименьшее число $m \in \mathbb{Q}$, для которого mP есть нулевой элемент группы, т. е. бесконечно удаленная точка, называется *порядком* точки P.

Задачи.

1. Докажите что точка $P = (0,2)$ на кривой $y^2 = x^3 + 4$ имеет порядок 3.

2. Докажите что точка $P = (2,4)$ на кривой $y^2 = x^3 + 4x$ имеет порядок 4.

3. Каждая из следующих точек имеет конечный порядок на соответствующей кривой. Найдите его.

а) $P = (0,4)$ на $y^2 = 4x^3 + 16$.
б) $P = (2,8)$ на $y^2 = 4x^3 + 16x$.
в) $P = (2,3)$ на $y^2 = x^3 + 1$.
г) $P = (3,8)$ на $y^2 = x^3 - 43x + 166$.
д) $P = (3,12)$ на $y^2 = x^3 - 14x^2 + 81x$.
е) $P = (0,0)$ на $y^2 + y = x^3 - x^2$.
ж) $P = (1,0)$ на $y^2 + xy + y = x^3 - x^2 - 3x + 3$.

4. Пусть f_n — функции на кривой

$$y^2 = x^3 + ax + b, \qquad 4a^3 + 27b^2 \neq 0,$$

задаваемые соотношениями

$$f_1 = 1,$$
$$f_2 = 2y,$$
$$f_3 = 3x^4 + 6ax^2 + 12bx - a^2,$$
$$f_4 = 4y(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3),$$
$$f_{2m} = 2f_m(f_{m+2}f_{m-1} - f_{m+2}^2 f_{m+1}^2), \ m \geqslant 3,$$
$$f_{2m+1} = f_{m+2}f_m^3 - f_{m-1}f_{m+1}^3, \ m \geqslant 2.$$

Пусть, далее, $g_n = xf_n^2 - f_{n-1}f_{n+1}$ и $4yh_n = f_{n+2}f_{n-1}^2 - f_{n-1}f_{n+1}^2$. Докажите, что

$$n(x,y) = (g_n/f_n^2, h_n/f_n^3).$$

§ 3. Некоторые примеры

В этом параграфе мы рассмотрим несколько примеров неособых кубических кривых, заданных уравнением в нормальной форме

$$y^2 = x^3 + ax + b$$

или же уравнением

$$y^2 + 2cy = x^3 + ax + b,$$

которое сводится к предыдущему заменой y на $y + c$.

Нам придется сменить обозначения. Дело в том, что в вопросах, связанных с рациональными точками на эллиптической кривой и редукцией по модулю p, обозначение кривой через E является традиционным. Поэтому до конца главы бесконечно удаленная точка кривой или единица группового закона будет упоминаться без какого-либо обозначения, а эллиптическая кривая будет обозначаться через E.

Нас будет интересовать в основном множество рациональных точек кривой E, которое мы будем обозначать $E(\mathbb{Q})$. Это множество, как уже говорилось, является абелевой группой, нулевым элементом которой служит бесконечно удаленная точка

кривой. В связи с этим бесконечно удаленную точку кривой тоже удобно считать рациональной точкой.

Пример 1. Рассмотрим следующую задачу: представить произведение двух последовательных целых чисел $y(y+1)$ в виде произведения трех последовательных целых чисел $(x-1)x(x+1) = x^3 - x$. Эта задача приводит к кривой E, задаваемой уравнением

$$y^2 + y = x^3 - x. \tag{3.1}$$

На этой кривой есть шесть очевидных точек с целочисленными

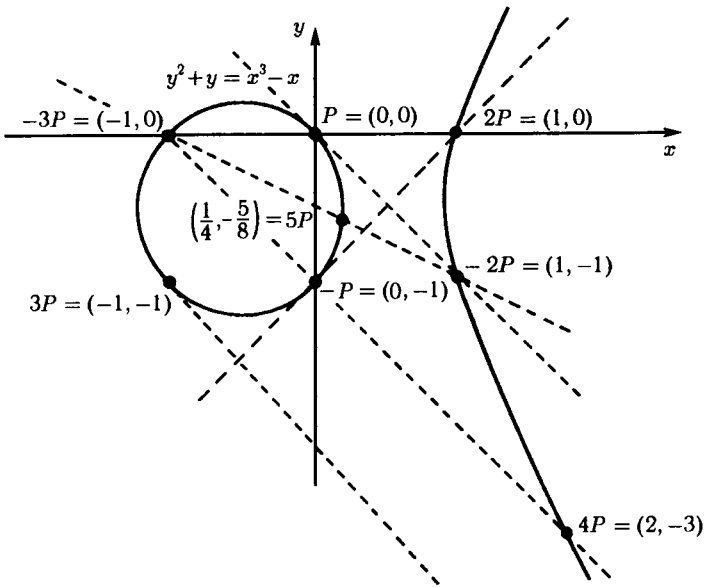


Рис. 38

координатами: $(0,0)$, $(1,0)$, $(-1,0)$, $(0,-1)$, $(1,-1)$ и $(-1,-1)$ (рис. 38).

Положим $P = (0,0)$. Тогда все указанные точки порождаются точкой P:

$$\begin{aligned} 3P &= (-1,-1),\\ 2P &= (1,0),\\ -P &= (0,-1),\\ -2P &= (1,-1),\\ -3P &= (-1,0). \end{aligned}$$

Точка P порождает бесконечную циклическую группу. Все точки вида $(2n+1)P$ лежат на замкнутой компоненте кривой, содержащей точку P. Точки вида $2nP$ расположены на некомпактной компоненте и при увеличении n стремятся к бесконечности.

Пример 2. Кривая E, задаваемая уравнением

$$y^2 + y = x^3 - x^2, \tag{3.2}$$

имеет четыре очевидные точки с целочисленными координатами: $(1,0)$, $(0,0)$, $(0,-1) = -(0,0)$ и $(1,-1) = -(1,0)$ (рис. 39).

Касательная к кривой E в точке $(1,0)$ пересекает кривую в точке $(0,-1)$. Это означает, что $2(1,0) = (0,0)$ и, стало быть, $2(1,-1) = (0,-1)$. Касательная к кривой E в точке $(0,0)$ пересекает ее в точке $(1,0)$, т. е. $2(0,0) = (1,-1)$. Из равенств

$$2(1,0) = (0,0),$$

$$2(0,0) = (1,-1) = -(1,0)$$

получаем

$$4(1,0) = -(1,0),$$

т. е. $5(1,0) = 0$. Следовательно, подмножество

$$\{0, (1,0), (0,0), (0,-1), (1,-1)\}$$

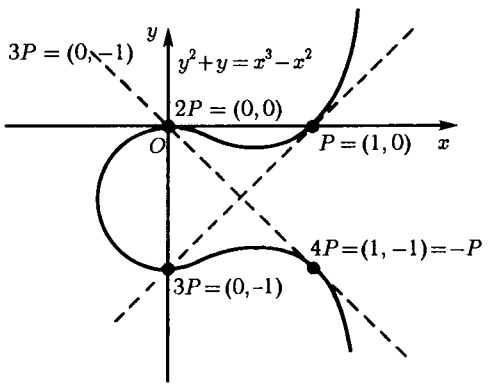


Рис. 39

образует в $E(\mathbb{Q})$ циклическую подгруппу порядка 5. Используя более развитую технику, можно показать, что других рациональных точек на этой кривой нет.

Пример 3. Рассмотрим кривую E, заданную уравнением

$$y^2 + y = x^3 + x^2. \tag{3.3}$$

На этой кривой есть четыре очевидные точки с целочисленными координатами: $(0,0)$, $(-1,0)$, $(0,-1) = -(0,0)$ и $(-1,-1) = -(-1,0)$ (рис. 40). Точка $P = (0,0)$ порождает бесконечную циклическую подгруппу в $E(\mathbb{Q})$. Например, нетрудно вычислить,

что

$$-3P = (1,1), \quad 2P = (-1,-1), \quad 3P = (1,-2),$$
$$4P = (2,3), \quad 5P = (-3/4,-9/8).$$

Касательная T к кривой в точке $-2P$ пересекает эту кривую в точке $4P$. Прямая L, проходящая через $2P$ и P, пересекает E

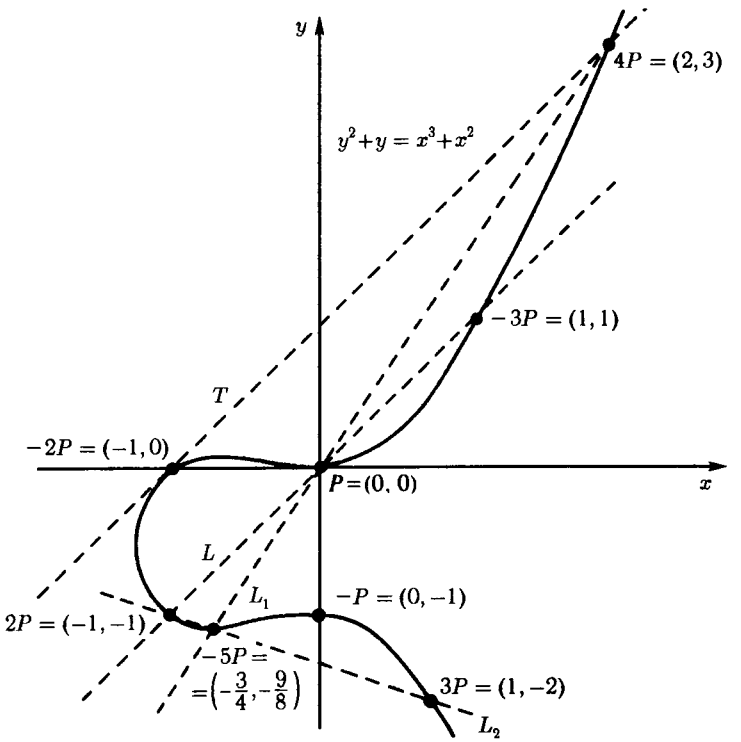


Рис. 40

в точке $-3P$. Точка $-5P$ вычисляется либо с помощью прямой L_1, проходящей через P и $4P$, либо с помощью прямой L_2, проходящей через точки $2P$ и $3P$.

Пример 4. Рассмотрим уравнение Ферма

$$u^3 + v^3 = w^3.$$

Его ненулевыми целочисленными решениями являются тройки

$(u,v,w) = (1,-1,0)$, $(1,0,1)$ и $(0,1,1)$. Все остальные целочисленные решения им пропорциональны. Кубическое уравнение $u^3+v^3 = w^3$ записано не в нормальной форме, однако с помощью преобразования

$$x = \frac{3w}{u+v}, \quad y = \frac{9}{2}\left(\frac{u-v}{u+v}\right) + \frac{1}{2}$$

мы получим кривую E, задаваемую уравнением в нормальной форме

$$y^2 - y = x^3 - 7. \tag{3.4}$$

Указанное преобразование переводит точку $(1,-1,0)$ в бесконечно удаленную. Точка $(1,0,1)$ переходит при этом в точку $(3,5)$, а точка $(0,1,1)$ — в $(3,-4)$. Таким образом, $E(\mathbb{Q})$ является циклической группой состоящей из трех элементов $\{0,(3,5),(3,-4)\}$, поскольку если бы $E(\mathbb{Q})$ содержало бы другие рациональные точки (x,y), то им соответствовали бы нетривиальные решения (u,v,w) уравнения Ферма $u^3+v^3=w^3$.

Убедимся, что $3P = 0$ для $P = (3,5)$, вычислив $-2P$ с помощью касательной к E в точке P. Из уравнения для производной $(2y-1)\,y' = 3x^2$ находим угловой коэффициент касательной к E в точке $(3,5)$:

$$\lambda = \frac{3x^2}{2y-1} = \frac{27}{9} = 3.$$

Следовательно, касательная задается уравнением

$$y = 5 + 3\,(x-3) = 3x - 4.$$

Она пересекает кривую (3.4) в точках, абсциссы которых удовлетворяют уравнению

$$(3x-4)^2 - (3x-4) = x^3 - 7,$$

т. е.

$$x^3 - 9x^2 + 27x - 27 = (x-3)^3 = 0.$$

Таким образом, абсцисса точки $-2P$ есть $x = 3$. Значит, $-2P = P$, откуда $3P = 0$.

Пример 5. Рассмотрим кривую

$$y^2 = x^3 + k, \tag{3.5}$$

где k — целое число. Соответствующее диофантово уравнение впервые было рассмотрено в XVII в. Ферма и Баше де Мезирьяком в частном случае $k = -2$ и впоследствии интенсивно изучалось. До сих пор неизвестно, при каких целых значениях k уравнение (3.5) имеет по крайней мере одно рациональное решение. Баше утверждал, но не доказал этого, что если есть рациональное решение (x, y), $xy \neq 0$, то метод касательных приводит к бесконечному числу рациональных решений. На современном языке это означает, что если группа $E(\mathbb{Q})$ рациональных точек кривой (3.5) ненулевая, то она бесконечная. С несколькими исключениями это утверждение было доказано в 1930 г. немецким математиком Фуэтером. В 1966 г. замечательно короткое доказательство результата Фуэтера получил английский математик Морделл (см. задачу 2 в конце этого параграфа).

Пример 6. В 1643 г. Ферма в письме к Мерсенну предложил найти такие целочисленные пифагоровы тройки (X, Y, Z) со взаимно простыми X, Y, Z, для которых гипотенуза Z и сумма катетов $X+Y$ являлись бы полными квадратами. Ответ, которым располагал Ферма, имеет вид

$$\begin{aligned} X &= 1061652293520, \\ Y &= 4565486027761, \\ Z &= 4687298610289, \end{aligned} \tag{3.6}$$

причем это наименьшее положительное решение.

Записывая эту задачу алгебраически, мы получаем систему уравнений

$$\begin{cases} x^2 + y^2 = Z^2, \\ Z = b^2, \\ X + Y = a^2. \end{cases}$$

Положим $e = X - Y$. Тогда

$$X = \frac{1}{2}(a^2 + e), \quad Y = \frac{1}{2}(a^2 - e).$$

Таким образом

$$b^4 = Z^2 = X^2 + Y^2 = \frac{1}{2}(a^4 + e^2)$$

и

$$e^2 = 2b^4 - a^4. \tag{3.7}$$

Поделив это соотношение на a^4, мы приходим к кривой

$$v^2 = 2u^4 - 1. \tag{3.8}$$

Если пара чисел (u, v) удовлетворяет уравнению (3.8) и если (x, y) заданы формулами

$$\begin{aligned} x &= \frac{2(v + 2u^2 - 1)}{(u-1)^2}, \\ y &= \frac{4((2u-1)v + 2u^3 - 1)}{(u-1)^3}, \end{aligned} \tag{3.9}$$

то имеет место соотношение

$$y^2 = x^3 + 8x. \tag{3.10}$$

Обратно, если точка (x, y) лежит на кривой (3.10), то пара (u, v), удовлетворяющая уравнению (3.8), находится из формул

$$\begin{aligned} u &= \frac{y - 2x - 8}{y - 4x + 8}, \\ v &= \frac{y^2 - 24x^2 + 48y - 16x - 64}{(y - 4x + 8)^2}. \end{aligned} \tag{3.11}$$

Например, точки $(x, y) = (0, 0)$, $(1, 3)$ и $(1, -3)$ соответствуют точкам $(u, v) = (-1, -1)$, $(-1, 1)$ и $(-13, -239)$. Заметим, что соответствие (3.9), (3.11) не является взаимно однозначным: точка $(u, v) = (13, 239)$ переходит при этом соответствии в $(x, y) = (8, 24)$, однако точка $(x, y) = (8, 24)$ отображается в особую точку. Точно так же, точка $(x, y) = (8, -24)$ переходит в $(u, v) = (1, -1)$, но $(u, v) = (1, -1)$ отображается в особую точку.

Итак, мы получаем, что рациональная точка на эллиптической кривой (3.10) задает рациональное решение (u, v) уравнения (3.8) и, следовательно, целочисленное решение (a, b, e) системы (3.7) со взаимно простыми a и b. Легко видеть, что числа a, b и e должны быть нечетными. Таким образом, мы приходим к пифагоровой тройке (X, Y, Z)

$$X = \frac{1}{2}(a^2 + e), \quad Y = \frac{1}{2}(a^2 - e), \quad Z = b^2 \tag{3.12}$$

и все, что остается сделать для решения задачи Ферма — это проверить, будут ли числа X и Y положительными. К сожалению, малые решения, такие как $(a, b, c) = (1, 13, 239)$, приводят к отрицательному Y.

Задачи

1. Пусть $P_0 = (x_0, y_0)$, $y_0 \neq 0$ — рациональная точка кривой

$$y^2 = x^3 + k.$$

Показать, что касательная в точке P_0 пересекает эту кривую в точке $P_1 = (x_1, y_1)$, где

$$x_1 = \frac{9x_0^4 - 8x_0y^2}{4y_0^2}, \quad y_1 = \frac{27x_0^6 - 36x_0^3y_0^2 + 8y_0^4}{8y_0^3}.$$

Показать, в частности, что кривая $y^2 = x^3 - 2$ имеет рациональные точки $P_0 = (3,5)$ и $P_1 = \left(\frac{129}{100}, \frac{383}{1000}\right)$.

2. (Морделл [B13, б]) Пусть k — целое число, свободное от шестых степеней и отличное от 1 и -432. Пусть, далее, кривая

$$y^2 = x^3 + k$$

имеет рациональную точку $P_0 = (x_0, y_0)$, причем $x_0y_0 \neq 0$.

а) Положив $x_0 = p/q^2$ и $y_0 = r/q^3$, где $(p,q,r) = 1$, $(p,q) = 1$ и $(q,r) = 1$, показать, что точка $P_1 = (x_1, y_1)$ из задачи 1 имеет координату x_1, задаваемую соотношением

$$q^2x_1 = \frac{9p^4}{4r^2} - 2p.$$

б) Доказать, что если $3p^2/2r$ не является целым числом, то $P_1 \neq P_0$.

в) Доказать, что если $3p^2/2r$ — целое число, то, применяя к точке P_1 указанный в предыдущей задаче процесс, можно прийти к рациональной точке $P_1' \neq P_0$.

г) Используя результаты задач а) — в), доказать, что кривая $y^2 = x^3 + k$ имеет бесконечно много рациональных точек.

3. а) Доказать что уравнение $y^2 = x^3 + 1$ не имеет рациональных решений, отличных от $(-1,0)$, $(0,\pm1)$, $(2,\pm3)$.

б) Доказать, что уравнение $y^2 = x^3 - 432$ не имеет рациональных решений, отличных от $(12,\pm36)$.

§ 4. Теорема Морделла

В 1901 г. великий французский математик А. Пуанкаре высказал предположение [A8], что все рациональные точки $E(\mathbb{Q})$ эллиптической кривой могут быть получены из некоторого их конечного числа с помощью операции сложения точек. В алгебраических терминах это утверждение можно сформулировать следующим образом.

Теорема 1. *На эллиптической кривой E, заданной уравнением с рациональными коэффициентами, группа $E(\mathbb{Q})$ рациональных точек является конечнопорожденной абелевой группой.*

Сам Пуанкаре считал это утверждение очевидным. Строгое доказательство гипотезы Пуанкаре впервые получил англичанин Л. Морделл в 1922 г. [B13, а]. За семь десятилетий появились различные обобщения и новые варианты доказательства этой теоремы. Мы приведем здесь один из них с небольшим пробелом (чтобы его заполнить, необходимы сведения из теории делимости в кольцах целых алгебраических чисел).

Итак, рассмотрим кривую

$$y^2 = x^3 + ax + b,$$

где $a, b \in \mathbb{Q}$ и $\Delta = -(4a^3 + 27b^2) \neq 0$, можно считать, что a и b — не только рациональные, но и целые числа. В самом деле, пусть

$$a = \frac{p}{q}, \quad b = \frac{r}{q}$$

(дроби в этих выражениях не предполагаются несократимыми). Положим

$$y = \frac{y_1}{q^3}, \quad x = \frac{x_1}{q^2}.$$

Тогда исходное уравнение перепишется в виде $y_1^2 = x_1^3 + pq^3x_1 + rq^5$; коэффициенты этого многочлена целые.

Пусть $\alpha_1, \alpha_2, \alpha_3$ — корни уравнения

$$y^2 = x^3 + ax + b.$$

Поскольку, по предположению, $a, b \in \mathbb{Z}$, то $\alpha_1, \alpha_2, \alpha_3$ — целые алгебраические числа. Итак,

$$y^2 = (x - \alpha_1)(x - \alpha_2)(x - \alpha_3). \tag{4.1}$$

При исследовании этого уравнения можно следить лишь за координатой x, так как каждому значению x_0 соответствуют два значения $\pm y_0$. Рассмотрим в качестве примера эллиптическую кривую $y^2 = x^3 - 2$. На этой кривой лежит точка $P_0 = (3,5)$. В § 3 мы показали, что

$$2P_0 = \left(\frac{129}{100}, -\frac{383}{1000}\right).$$

При этом

$$x_1 - \alpha_1 = \frac{129}{100} - \sqrt[3]{2} = \left(\frac{9 - 6\sqrt[3]{2} - 2\sqrt[3]{4}}{10}\right)^2,$$

т. е. точка $2P_0 = (x_1, y_1)$ имеет следующее свойство: число $x_1 - \alpha_1$ является квадратом в поле $\mathbb{Q}(\sqrt[3]{2})$. Это свойство играет в доказательстве теоремы Морделла решающую роль.

Покажем, что если E — определенная над $\mathbb{Q}$ эллиптическая кривая (4.1) и α — одно из целых алгебраических чисел $\alpha_1, \alpha_2, \alpha_3$, то отображение $E(\mathbb{Q}) \to \mathbb{Q}(\alpha)$, при котором точке $(x_0, y_0) \in E(\mathbb{Q})$ ставится в соответствие элемент $x_0 - \alpha \in \mathbb{Q}(\alpha)$ может быть включено в коммутативную диаграмму

$$\begin{array}{ccc} E(\mathbb{Q}) & \longrightarrow & \mathbb{Q}(\alpha) \\ \downarrow & & \downarrow \\ E(\mathbb{Q})/2E(\mathbb{Q}) & \longrightarrow & \mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2, \end{array}$$

причем сложению точек в $E(\mathbb{Q})$ соответствует умножение смежных классов в $\mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2$.

Получим сначала явные формулы для сложения точек кривой $y^2 = x^3 + ax + b$ в более удобном для наших целей виде. Пусть прямая $y = px + q$ пересекает кривую $y^2 = x^3 + ax + b$ в трех точках (x_i, y_i), $i = 1,2,3$. Тогда

$$x^3 + ax + b - (px + q) = (x - x_1)(x - x_2)(x - x_3).$$

В частности, если $x = \alpha$ — один из корней многочлена $x^3 + ax + b$, то учитывая, что $\alpha^3 + a\alpha + b = 0$, получим

$$(x_1 - \alpha)(x_2 - \alpha)(x_3 - \alpha) = (p\alpha + q)^2.$$

Ясно также, что

$$p\alpha + q = \frac{y_1(x_2 - \alpha) - y_2(x_1 - \alpha)}{x_2 - x_1}.$$

Следовательно,

$$x_3 - \alpha = \frac{1}{(x_1 - \alpha)(x_2 - \alpha)}\left(\frac{y_1(x_2 - \alpha) - y_2(x_1 - \alpha)}{x_2 - x_1}\right)^2. \quad (4.2)$$

В случае $x_1 = x_2$ (т. е. для касательной) получаем

$$p = \frac{3x_1^2 + a}{2y_1}.$$

Поэтому

$$x_3 - \alpha = \left(\frac{1}{\alpha - x_1}\right)^2\left(\frac{(3x_1^2 + a)(\alpha - x_1)}{2y_1} + y_1\right)^2 =$$

$$= \left(\frac{1}{2y_1}\left(3x_1^2 + a + \frac{2y_1^2}{\alpha - x_1}\right)\right)^2. \quad (4.3)$$

Если воспользоваться тем, что

$$y_1^2 = (x_1 - \alpha)(x_2 - \alpha_2)(x_3 - \alpha_3), \quad \alpha + \alpha_2 + \alpha_3 = 0,$$
$$\alpha\alpha_2\alpha_3 = -b \quad \text{и} \quad \alpha^3 + a\alpha + b = 0,$$

то (4.3) можно привести к виду

$$x_3 - \alpha = \left(\frac{x_1^2 - a - 2\alpha x_1 - 2\alpha^2}{2y_1}\right)^2. \quad (4.4)$$

Теперь мы готовы доказать коммутативность изображенной выше диаграммы. В группе $\mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2$ все неединичные элементы имеют порядок 2. Поэтому формула (4.2) показывает, что если точкам P и Q соответствуют классы $A, B \in \mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2$, то точке $P+Q$ соответствует класс AB. Из формулы (4.4) вытекает, что точке $2P$ соответствует единичный класс E. Это утверждение можно обратить: если точка Q соответствует единичному классу, то $Q = 2P$. При доказательстве мы ограничимся лишь случаем, когда элементы $1, \alpha$ и α^2 линейно независимы над $\mathbb{Q}$. Пусть

$$x - \alpha = (u_0 + u_1\alpha + u_2\alpha^2)^2 =$$
$$= u_0^2 + 2u_0u_1\alpha + (u_1^2 + 2u_0u_1)\alpha^2 + 2u_1u_2\alpha^3 + u_2^2\alpha^4 =$$
$$= (u_0^2 - 2bu_1u_2) + (2u_0u_1 - 2au_1u_2 - bu_2^2) +$$
$$+ (u_1^2 + 2u_0u_2 - au_2^2)\alpha^2.$$

Тогда

$$\text{а)}\quad u_0^2 - 2bu_1u_2 = x,$$
$$\text{б)}\quad 2u_0u_1 - 2au_1u_2 - bu_2^2 = -1,$$
$$\text{в)}\quad u_1^2 + 2u_0u_2 - au_2^2 = 0.$$

Умножим б) и в) соответственно на $-u_2$ и u_1 и сложим полученные равенства. В результате получим

$$u_1^3 + au_1u_2^2 + bu_2^2 = u_2.$$

А так как $u_2 \neq 0$, то

$$\frac{1}{u_2^2} = \left(\frac{u_1}{u_2}\right)^3 + a\left(\frac{u_1}{u_2}\right) + b,$$

т. е. точка $P = (x', y')$, где $x' = \dfrac{u_1}{u_2}$ и $y' = \dfrac{1}{u_2}$, лежит на рассматриваемой кривой. Докажем, что $Q = 2P$. Из соотношения в) вытекает, что

$$x'^2 + 2u_0y' - a = 0,$$

поэтому $u_0 = \dfrac{a - x'^2}{2y'}$. Кроме того, $u_2 = \dfrac{1}{y'}$, $u_1 = x'u_2 = \dfrac{x'}{y'}$. Следовательно,

$$u_0 + u_1\alpha + u_2\alpha^2 = \frac{-x'^2 + a}{2y'} + \frac{x'}{y'}\alpha + \frac{1}{y'}\alpha^2 = -\frac{x'^2 - \alpha - 2\alpha x' - 2\alpha^2}{2y'}.$$

В силу (4.4) отсюда вытекает, что $Q = 2P$. Таким образом, если точки P и Q соответствуют одному классу, то $P+Q$ соответствует классу E, поэтому существует такая точка R, что $P+Q = 2R$. Этим завершается доказательство коммутативности нашей диаграммы.

Важно отметить, что образ группы $E(\mathbb{Q})$ в $\mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2$ конечен (мы опускаем доказательство, чтобы избежать довольно длинного экскурса в теорию делимости в кольцах целых алгебраических чисел). Отсюда следует, что рациональные точки нашей кривой попадают в конечное число классов $\mathbb{Q}(\alpha)/\mathbb{Q}(\alpha)^2$. Пусть представителями этих классов являются точки $Q_1, \ldots, Q_m$. Возьмем произвольную рациональную точку P_0 данной кривой.

Она попадает в один класс с некоторой точкой Q_{i_1}, и поэтому $P_0 + Q_{i_1} = 2P_1$. Аналогично, $P_1 + 2Q_{i_2} = 2P_2$. Поэтому

$$P_0 + Q_{i_1} + 2Q_{i_2} = 2(P_1 + Q_{i_2}) = 4P_2.$$

Продолжая аналогичные рассуждения, найдем, что

$$P_0 + Q_{i_1} + 2Q_{i_2} + \ldots + 2^kQ_{i_{k+1}} = 2^{k+1}P_{k+1}.$$

Ниже мы покажем, что для любой точки P_0 после конечного числа шагов получается точка P_{k+1}, числители и знаменатели координат которой ограничены константой C, не зависящей от P_0. Таких точек конечное число и, стало быть, любая точка P_0 принадлежит группе, порожденной точками $Q_1, \ldots, Q_m$ и еще некоторым конечным набором точек. А это и есть искомая теорема Морделла.

Итак, нам осталось доказать, что указанная выше процедура построения P_{k+1} приводит к конечному множеству точек.

Точку (x_0, y_0) кривой $y^2 = x^3 + ax + b$ можно представить в виде $(p/s^2, t/s^3)$, где p/s^2 и t/s^3 — несократимые дроби. В самом деле, пусть $x_0 = p/q$ — несократимая дробь. Тогда

$$(y_0q^2)^2 = q(p^3 + apq^2 + bq^3) = qr,$$

где $(q, r) = (q, p^3) = 1$. Поэтому $q = s^2$, $r = t^2$ и $y_0^2s^8 = s^2t^2$, т. е. $y = t/s^3$. Так как $(p, q) = 1$ и $(q, r) = 1$, то $(p, s) = 1$ и $(t, s) = 1$.

Для точки $P_0 = (x_0, y_0)$ положим $\lambda_0 = \max(|p|, s^2)$. Тогда

$$t^2 \leqslant \lambda_0^3(1 + |a| + |b|),$$

т. е.

$$|t| \leqslant c_1\lambda_0^{3/2}.$$

Аналогично, для точек Q_{i_1}, P_1 и $2P_1$ определим числа ρ, λ_1 и ω соответственно. Нас интересует оценка для λ_1 через λ_0. Оценим сначала ω. Для удобства введем следующие обозначения координат точек:

$$P_0 = \left(\frac{x}{z^2}, \frac{y}{z^3}\right), \quad Q_{i_1} = \left(\frac{p}{r^2}, \frac{q}{r^3}\right), \quad P_1 = \left(\frac{x_1}{z_1^2}, \frac{y_1}{z_1^3}\right), \quad 2P_1 = \left(\frac{s}{u^2}, \frac{t}{u^3}\right).$$

При этом числа p, q, r можно считать ограниченными константой (числа s, t, u используются лишь как промежуточный результат вычислений; окончательный результат — числа x_1, y_1, z_1).

Формулу сложения точек кубической кривой можно преобразовать к следующему виду:

$$x_3 = -x_1 - x_2 + \left(\frac{y_1 - y_2}{x_1 - x_2}\right)^2 = \frac{(x_1x_2 + a)(x_1 + x_2) + 2b - 2y_1y_2}{(x_1 - x_2)^2}.$$

Поэтому из равенства $P_0 + Q_{i_1} = 2P_1$ получаем

$$\frac{s}{u^2} = \frac{(px + ar^2z^2)(r^2x + pz^2) + 2br^4z^4 - 2qryz}{(r^2x - pz^2)^2}.$$

Числа x и z^2 имеют порядок λ_0, z^4 — порядок λ_0^2, а yz — порядок $\lambda_0^{3/2}\lambda_0^{1/2} = \lambda_0^2$. Поэтому $\omega \leqslant c_2\lambda_0^2$.

Покажем теперь, что

$$\lambda_1 \leqslant c_3\omega^{1/4} \leqslant c_4\lambda_0^{1/2}.$$

Тогда

$$\lambda_2 \leqslant c_4c_4^{1/2}\lambda_0^{1/4}$$

и поэтому $\lambda_n \leqslant c\lambda_0^{1/2^n}$, где $c = \max(1, c_4^2)$. Следовательно, для любого λ_0 при достаточно больших n число λ_n ограничено константой c.

Формулу (4.4) можно записать в виде

$$\left(\frac{s}{u^2} - \alpha\right)^{1/2} = \left(\frac{x_1^2}{z_1^4} - a - 2\alpha\frac{x_1}{z_1^2} - 2\alpha^2\right)\frac{z_1^3}{2y_1},$$

т. е.

$$(s - \alpha u^2)^{1/2} = \frac{u}{2y_1z_1}(x_1^2 - az_1^4 - 2\alpha x_1z_1^2 - 2\alpha^2z_1^4) = e_0 + e_1\alpha + e_2\alpha^2, \quad (4.5)$$

где

$$e_0 = \frac{u}{2y_1z_1}(x_1^2 - az_1^4), \quad e_1 = -\frac{ux_1z_1}{y_1}, \quad e_2 = \frac{uz_1^3}{y_1}.$$

Выражение в левой части есть целое алгебраическое число, поэтому $e_0 + e_1\alpha + e_2\alpha^2$ также целое алгебраическое число.

Лемма 2. *Пусть $e_0+e_1\alpha+e_2\alpha^2$ — целое алгебраическое число, причем $e_i \in \mathbb{Q}$ и $\alpha^3+a\alpha+b = 0$. Тогда $e_i\Delta \in \mathbb{Z}$, где Δ — дискриминант многочлена $x^3 + ax + b$.*

Доказательство. Пусть $\alpha_1, \alpha_2, \alpha_3$ — корни многочлена $x^3 + ax + b$ и $\sigma_1, \sigma_2, \sigma_3$ — автоморфизмы поля $\mathbb{C}$, переводящие α в $\alpha_1, \alpha_2, \alpha_3$ соответственно. Для $\gamma \in \mathbb{Q}(\alpha)$ определим след $\operatorname{tr}(\gamma)$, полагая

$$\operatorname{tr}(\gamma) = \sigma_1(\gamma) + \sigma_2(\gamma) + \sigma_3(\gamma).$$

При этом $\operatorname{tr}(\gamma) \in \mathbb{Q}$ и если γ — целое алгебраическое число, то $\operatorname{tr}(\gamma) \in \mathbb{Z}$.

Умножим равенство $\gamma = e_0 + e_1\alpha + e_2\alpha^2$ на α и α^2, а затем возьмем след:

$$\operatorname{tr}(\gamma) = 3e_0 + e_1\operatorname{tr}(\alpha) + e_2\operatorname{tr}(\alpha^2),$$
$$\operatorname{tr}(\alpha\gamma) = e_0\operatorname{tr}(\alpha) + e_1\operatorname{tr}(\alpha^2) + e_2\operatorname{tr}(\alpha^3),$$
$$\operatorname{tr}(\alpha^2\gamma) = e_0\operatorname{tr}(\alpha^2) + e_1\operatorname{tr}(\alpha^3) + e_2\operatorname{tr}(\alpha^4).$$

Эти три равенства можно рассмотреть как систему уравнений с целочисленными коэффициентами для e_0, e_1, e_2. Определитель этой системы равен

$$\begin{vmatrix} 3 & \alpha_1+\alpha_2+\alpha_3 & \alpha_1^2+\alpha_2^2+\alpha_3^2 \\ \alpha_1+\alpha_2+\alpha_3 & \alpha_1^2+\alpha_2^2+\alpha_3^2 & \alpha_1^3+\alpha_2^3+\alpha_3^3 \\ \alpha_1^2+\alpha_2^2+\alpha_3^2 & \alpha_1^3+\alpha_2^3+\alpha_3^3 & \alpha_1^4+\alpha_2^4+\alpha_3^4 \end{vmatrix} =$$
$$= \begin{vmatrix} 1 & 1 & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1^2 & \alpha_2^2 & \alpha_3^2 \end{vmatrix} \cdot \begin{vmatrix} 1 & \alpha_1 & \alpha_1^2 \\ 1 & \alpha_2 & \alpha_2^2 \\ 1 & \alpha_3 & \alpha_3^2 \end{vmatrix} =$$
$$= (\alpha_1 - \alpha_2)^2(\alpha_1 - \alpha_3)^2(\alpha_2 - \alpha_3)^2 = \Delta. \quad \square$$

Итак, числа $(2e_0 - ae_2)\Delta = \dfrac{u\Delta}{y_1z_1}x_1^2$ и $-e_2\Delta = \dfrac{u\Delta}{y_1z_1}z_1^4$ — целые. Так как $\dfrac{x_1}{z_1^2}$ и $\dfrac{y_1}{z_1^3}$ — несократимые дроби, то число $\dfrac{u\Delta}{y_1z_1}$ — также целое. В самом деле, если бы эта дробь была несократима и в ее знаменатель входил бы делитель y_1 (соответственно, z_1), то он не мог бы сократиться с z_1^4 (соответственно, x_1^2). Следовательно, x_1^2 и z_1^4 являются делителями чисел $(2e_0 - ae_2)\Delta$ и $-e_2\Delta$ соответственно. Написав равенство (4.5) для чисел α_1, α_2 и α_3, мы получим, что $(2e_0 - ae_1)\Delta$ и $e_2\Delta$ линейно выражаются через $(s - \alpha_iu^2)^{1/2}$, $i = 1, 2, 3$, а значит, имеют порядок $\omega^{1/2}$. Итак, мы

показали, что $\lambda_1 \leqslant c_1 \lambda_0^{1/2}$ и, тем самым, установили конечность множества точек P_{k+1}, что и требовалось.

§ 5. Ранг и группа кручения эллиптической кривой

Согласно результатам предыдущего параграфа группа $E(\mathbb{Q})$ рациональных точек эллиптической кривой, определенной над полем $\mathbb{Q}$, является конечнопорожденной абелевой группой. Как любая конечнопорожденная абелева группа, $E(\mathbb{Q})$ допускает следующее разложение:

$$E(\mathbb{Q}) = \mathbb{Z}^{r_E} \times \operatorname{Tors} E(\mathbb{Q}),$$

где r_E — ранг группы $E(\mathbb{Q})$, а $\operatorname{Tors} E(\mathbb{Q})$ — подгруппа элементов конечного порядка в $E(\mathbb{Q})$. Число r_E называется *рангом эллиптической кривой* E, а подгруппа $\operatorname{Tors} E(\mathbb{Q})$ — *группой кручения* этой эллиптической кривой. Можно показать, что ранг эллиптической кривой сохраняется при бирациональных преобразованиях.

Ранг r_E вычислен для многих эллиптических кривых над $\mathbb{Q}$. В большинстве случаев он очень мал: чаще всего он равен 0, 1, 2 или 3. В качестве иллюстрации, в табл. 1 и 2 мы приведем значения рангов кривых $y^2 = x^3 + ax$ и $y^2 = x^3 + a$ для некоторых значений a [Б31].

Таб л. 1.

Ранги эллиптических кривых E, задаваемых уравнениями $y^2 = x^3 + ax$

Ранг	Значение параметра a
$r_E = 0$	$a = 1,\ 2,\ 4,\ 6,\ 7,\ 10,\ 11,\ 12,\ 22,$ $-1,\ -3,\ -4,\ -8,\ -9,\ -11,\ -13,\ -18,\ -19$
$r_E = 1$	$a = 3,\ 5,\ 8,\ 9,\ 13,\ 15,\ 18,\ 19,\ 20,$ $-2,\ -5,\ -6,\ -7,\ -10,\ -12,\ -14,\ -15,\ -20$
$r_E = 2$	$a = 14,\ 33,\ 34,\ 39,\ 46,\ -17,\ -56,\ -65,\ -77$
$r_E = 3$	$a = -82$

Таб л. 2.

Ранги эллиптических кривых E, задаваемых уравнениями $y^2 = x^3 + a$

Ранг	Значение параметра a
$r_E = 0$	$a = 1,\ 4,\ 6,\ 7,\ 13,\ 14,\ 16,\ 20,\ 21,$ $-1,\ -3,\ -5,\ -6,\ -8,\ -9,\ -10,\ -14,\ -432$
$r_E = 1$	$a = 2,\ 3,\ 5,\ 8,\ 9,\ 10,\ 11,\ 12,\ 18,$ $-2,\ -4,\ -7,\ -13,\ -15,\ -18,\ -19,\ -20,\ -21$
$r_E = 2$	$a = 15,\ 17,\ 24,\ 37,\ 43,\ -11,\ -26,\ -39,\ -47$
$r_E = 3$	$a = 113,\ 141,\ 316,\ 346,\ 359,\ -174,\ -307,\ -362$

До сих пор неизвестно, существуют ли эллиптические кривые сколь угодно большого ранга. В 1986 г. французскому математику Местру удалось построить примеры эллиптических кривых всех рангов от 3 до 14 включительно. Например, кривая

$$y^2 + 9767y = x^3 + 3576x^2 + 425x - 2412$$

имеет ранг $r_E \geqslant 9$, а кривая

$$y^2 + 357573631y = x^3 + 2597055x^2 - 549082x - 19608054$$

имеет ранг $r_E \geqslant 14$.

Одна из наиболее знаменитых гипотез в современной теории чисел связывает число r_E с порядком нуля в $s = 1$ некоторой аналитической функции, соответствующей кривой E. Эта гипотеза была сформулирована в 1965 г. английскими математиками Бёрчем и Суиннертон-Дайером. Перед тем, как обратится к ней, нам нужно ввести некоторые новые понятия.

Пусть $y^2 = x^3 + ax + b$ — некоторая эллиптическая кривая, причем

$$\Delta = -(4a^3 + 27b^2) \neq 0.$$

Как уже указывалось в предыдущем параграфе, без ограничения общности мы можем считать, что a и b — целые числа.

Пусть p — некоторое простое число. Рассмотрим сравнение

$$y^2 \equiv x^3 + ax + b \pmod{p},$$

или, что эквивалентно, уравнение

$$y^2 = x^3 + \bar{a}x + \bar{b}, \quad \bar{a}, \bar{b} \in \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p. \tag{5.1}$$

Если простое число p не делит дискриминант Δ, то уравнение (5.1) определяет эллиптическую кривую E_p над полем $\mathbb{F}_p$. В этом случае говорят, что кривая E имеет *хорошую редукцию*. Кривую E_p называют *редукцией кривой E по модулю p*. Обозначим через N_p число точек кривой E_p с координатами в $\mathbb{F}_p$, включая бесконечно удаленную точку.

Например, уравнение $y^2 = x^3 + 3x$, рассматриваемое над полем $\mathbb{F}_5$, имеет решения

$$(0,0),\ (1,\pm 2),\ (2,\pm 2),\ (3,\pm 1),\ (4,\pm 1),\ \infty$$

и поэтому $N_5 = 10$.

Что можно сказать о числе N_p для произвольной кривой $y^2 = x^3 + \bar{a}x + \bar{b}$ над полем $\mathbb{F}_p$? Любая такая кривая содержит по меньшей мере бесконечно удаленную точку. С другой стороны, каждый элемент x из $\mathbb{F}_p$ дает не более двух значений y, поэтому, с учетом бесконечно удаленной точки, число N_p не превышает $2p+1$. Таким образом, имеют место очевидные неравенства

$$1 \leqslant N_p \leqslant 2p + 1.$$

Эти неравенства можно переписать в следующем виде:

$$|p + 1 - N_p| \leqslant p.$$

В 1934 г. немецкий математик Г. Хассе получил более точную оценку [B20]:

$$|p + 1 - N_p| \leqslant 2\sqrt{p}.$$

Отправляясь от чисел N_p, определим так называемую *L-функцию* рациональной эллиптической кривой E, полагая

$$L(E,s) = \prod_{p|\Delta} \left(\frac{1}{1 - a_p p^{-s}}\right) \prod_{p\nmid\Delta} \left(\frac{1}{1 - a_p p^{-s} + p^{1-2s}}\right), \tag{5.2}$$

где $a_p = p + 1 - N_p$ и Δ — дискриминант данной кривой.

Из оценки Хассе можно без труда показать, что указанное бесконечное произведение сходится при $\operatorname{Re} s > 3/2$. Существует гипотеза, что функция $L(E,s)$ может быть аналитически продолжена на всю комплексную плоскость.

Гипотеза 1 (Хассе — Вейль). *Для любой эллиптической кривой E, определенной над $\mathbb{Q}$, существует такое положительное целое число N и знак $\varepsilon = \pm 1$, что модифицированная L-функция*

$$\Lambda(E,s) = N^{s/2}(2\pi)^{-s}\,\Gamma(s)\,L(E,s),$$

где через $\Gamma(s)$ обозначена Γ-функция Эйлера, удовлетворяет функциональному уравнению

$$\Lambda(E,s) = -\varepsilon\,\Lambda(E,2-s).$$

Несмотря на огромные усилия многих первоклассных математиков, эта гипотеза до лета 1993 г. была доказана лишь в некоторых частных случаях. В июне 1993 г. произошло знаменательное событие — американский математик А. Уайлс анонсировал доказательство гипотезы Таниямы для полустабильных эллиптических кривых, из которой, в свою очередь, вытекает гипотеза Хассе — Вейля. Мы рассмотрим гипотезу Таниямы в следующем параграфе в связи с доказательством последней теоремы Ферма.

Используя обширный эмпирический материал о кривых вида $y^2 = x^3 + ax$ и $y^2 = x^3 + a$ и предположив, что $L(E,s)$ может быть продолжена на всю комплексную плоскость, Бёрч и Суиннертон-Дайер пришли к следующей гипотезе.

Гипотеза 2 (Бёрч, Суиннертон-Дайер). *Пусть E — определенная над $\mathbb{Q}$ эллиптическая кривая. Тогда ее ранг r_E равен порядку нуля функции*

$$\widetilde{L}(E,s) = \prod_{p\nmid\Delta} \frac{1}{1 - a_p p^{-s} + p^{1-2s}} \tag{5.3}$$

в точке $s = 1$.

Это предположение, так же как и гипотеза Хассе — Вейля, вытекает из гипотезы Таниямы.

Займемся теперь группой кручения $\operatorname{Tors} E(\mathbb{Q})$. Чтобы проиллюстрировать встречающиеся здесь возможности, вычислим

группу кручения для семейства кривых $y^2 = x^3 + ax$. Без ограничения общности мы можем считать, что a — ненулевое целое число, свободное от четвертых степеней. В самом деле, при подстановке $x \mapsto q^2x$, $y \mapsto q^3y$ уравнение $y^2 = x^3 + ax$ преобразуется к виду

$$q^6y^2 = q^6\left(x^3 + \left(\frac{a}{q^4}\right)x\right).$$

Теорема 3. *Пусть E — эллиптическая кривая, задаваемая уравнением $y^2 = x^3 + ax$, где a — свободное от четвертых степеней целое число. Тогда*

$$\operatorname{Tors} E(\mathbb{Q}) = \begin{cases} \mathbb{Z}/2 \oplus \mathbb{Z}/2, & \text{если } -a = m^2,\ m \in \mathbb{N},\ a \neq 4; \\ \mathbb{Z}/4, & \text{если } a = 4; \\ \mathbb{Z}/2, & \text{в остальных случаях.} \end{cases}$$

Доказательство. Найдем среди точек $E(\mathbb{Q})$ точки порядка 2. Все они имеют вид $(x, 0)$, где x — корень кубического уравнения $x^3 + ax = 0$. Таким образом, три точки порядка 2 существуют тогда и только тогда, когда $-a = m^2$, $m \in \mathbb{N}$. При этом точка $(0, 0)$ всегда является точкой второго порядка.

Обратимся теперь к точкам четвертого порядка. Пусть (x_0, y_0) одна из таких точек. Тогда $2(x_0, y_0)$ — точка второго порядка. Имеется две возможности. Либо $2(x_0, y_0) = (0, 0)$, либо $2(x_0, y_0) = (\pm m, 0)$.

В первом случае существует прямая L, задаваемая уравнением $y = \lambda x$, проходящая через $(0, 0)$ и касающаяся кривой E в точке (x, y). Поэтому уравнение

$$(\lambda x)^2 = x^3 + ax, \tag{5.3}$$

имеет однократный корень 0 и двукратный x_0. Уравнение (5.3) преобразуется к виду $x(x^2 - \lambda^2 x + a) = 0$. Так как прямая L касается E в точке (x_0, y_0), то квадратное уравнение $x^2 - \lambda^2 x + a = 0$ имеет двукратный корень x_0, т. е. его дискриминант $\lambda^4 - 4a$ равен нулю. Поскольку число a свободно от четвертых степеней, уравнение $\lambda^4 - 4a = 0$ имеет рациональное решение только при $a = 4$. В этом случае точками четвертого порядка, удовлетворяющими уравнению $2(x, y) = (0, 0)$, будут точки $(2, 4)$ и $(2, -4)$.

Во втором случае, который имеет место только тогда, когда выполнено равенство $-a = m^2$, $m \in \mathbb{N}$, существует прямая L,

задаваемая уравнением $y = \lambda(x \pm m)$, проходящая через точку $(\mp m, 0)$ и касающаяся кривой E в точке (x_0, y_0). Уравнение $(\lambda(x \pm m))^2 = x^3 + ax$ преобразуется к виду $(x \pm m)(x^2 - (\lambda^2 \pm m)x \mp m\lambda^2) = 0$. Так как прямая L касается кривой E, то дискриминант квадратного трехчлена $x^2 - (\lambda^2 \pm m)x \mp m\lambda^2$ равен нулю. То есть

$$\lambda^4 \pm 6\lambda^2 m + m^2 = 0. \tag{5.4}$$

Из полученного уравнения следует, что $m = \lambda^2(\pm 3 \pm 2\sqrt{2})$. В силу этого, соотношение (5.4) не может быть выполнено при $m \in \mathbb{N}$ и $\lambda \in \mathbb{Q}$.

При помощи аналогичных рассуждений можно показать, что на кривой E при $a = 4$ не существует рациональных точек, удовлетворяющих уравнению $2(x, y) = (2, \pm 4)$.

Отсюда следует, что точки с порядками равными 2^k, $k \in \mathbb{N}$, образуют, в зависимости от числа a, подгруппы $\mathbb{Z}/2 \oplus \mathbb{Z}/2$, $\mathbb{Z}/4$ и $\mathbb{Z}/2$. Таким образом, для доказательства теоремы остается показать, что в $E(\mathbb{Q})$ не существует точек нечетного порядка.

Доказательство этого факта мы приведем от противного. Предположим, что в $E(\mathbb{Q})$ имеется точка $P \neq 0$, для которой $3P = 0$, т. е. $2P = -P$. Тогда касательная прямая $y = \lambda x + \beta$ к кривой E в точке P, будучи подставленной в уравнение $y^2 = x^3 + ax$ должна давать полный куб, т. е. должно выполняться уравнение

$$0 = x^3 + ax - (\lambda x + \beta)^2 = (x - r)^3, \tag{5.5}$$

где r — первая координата точки P. Раскрывая скобки в соотношении (5.5), получаем

$$0 = x^3 - \lambda^2 x^2 + (a - 2\beta\lambda)x - \beta^2 = x^3 - 3rx^2 + 3r^2x - r^3,$$

откуда $3r = \lambda^2$, $r^3 = \beta^2$, что дает соотношение $\beta^2 = \lambda^6/27$. Кроме того, третье соотношение между коэффициентами

$$3r^2 = a - 2\beta\lambda$$

приводит к равенству

$$3\left(\frac{\lambda^4}{9}\right) = a - 2\left(\frac{\lambda^4}{3\sqrt{3}}\right),$$

невозможному при рациональных a и λ. □

Подобным образом, можно показать, что если E — эллиптическая кривая, задаваемая уравнением $y^2 = x^3 + a$, где a — свободное от шестых степеней число, то

$$\operatorname{Tors} E(\mathbb{Q}) = \begin{cases} \mathbb{Z}/6\mathbb{Z}, & \text{если } a = 1; \\ \mathbb{Z}/3\mathbb{Z}, & \text{если } a = \pm m^2,\ m \in \mathbb{N},\ a \neq 1,\ a \neq -432; \\ \mathbb{Z}/2, & \text{если } a = \pm m^3,\ m \in \mathbb{Z},\ a \neq 1; \\ 0, & \text{в остальных случаях.} \end{cases}$$

Структуру группы $\operatorname{Tors} E(\mathbb{Q})$ для кривых в форме Вейерштрасса $y^2 = f(x)$ с целыми коэффициентами во многом проясняет следующее утверждение.

Теорема 4. *Пусть E — эллиптическая кривая $y^2 = f(x)$, где*

$$f(x) = x^3 + ax^2 + bx + c$$

— многочлен третьей степени с целыми коэффициентами. Если (x, y) — точка конечного порядка в E, то $x, y \in \mathbb{Z}$ и y делит дискриминант Δ многочлена $f(x)$.

Из этой теоремы вытекает эффективный вычислительный алгоритм для нахождения группы $\operatorname{Tors} E(\mathbb{Q})$ эллиптической кривой

$$y^2 = x^3 + ax^2 + bx + c, \quad a, b, c \in \mathbb{Z}.$$

Рассмотрим конечное множество всех делителей y_0 дискриминанта $\Delta = 18a^2b^2 - 4a^3c + 18abc - 4b^3 - 27c^2$ и найдем все целочисленные решения x_0 кубического уравнения

$$x^3 + ax^2 + bx + c - y_0^2 = 0.$$

Тогда все точки конечного порядка, не равные нулю, содержатся среди найденного множества точек $\{(x_0, y_0)\}$.

В 1976 г. американский математик Б. Мазур [B11] доказал удивительную теорему о структуре групп кручения эллиптических кривых.

Теорема 5. *Пусть E — произвольная эллиптическая кривая, определенная над полем рациональных чисел $\mathbb{Q}$. Тогда $\operatorname{Tors} E(\mathbb{Q})$ изоморфна одной из следующих пятнадцати групп: $\mathbb{Z}/m\mathbb{Z}$ при $m \leqslant 10$ или $m = 12$, $\mathbb{Z}/2\mathbb{Z} \oplus \oplus \mathbb{Z}/2m\mathbb{Z}$ при $m \leqslant 4$.*

В следующей ниже таблице приведены примеры рациональных эллиптических кривых, для которых реализуются эти группы кручения.

Таб л. 3. Примеры групп кручения эллиптических кривых

Эллиптическая кривая E	Группа $\operatorname{Tors} E(\mathbb{Q})$
$y^2 = x^3 + 2$	0
$y^2 = x^3 + 8$	$\mathbb{Z}/2$
$y^2 = x^3 + 4$	$\mathbb{Z}/3$
$y^2 = x^3 + 4x$	$\mathbb{Z}/4$
$y^2 + y = x^3 - x^2$	$\mathbb{Z}/5$
$y^2 = x^3 + 1$	$\mathbb{Z}/6$
$y^2 - xy + 2y = x^3 + 2x^2$	$\mathbb{Z}/7$
$y^2 + 7xy - 6y = x^3 - 6x^2$	$\mathbb{Z}/8$
$y^2 + 3xy + 6y = x^3 + 6x^2$	$\mathbb{Z}/9$
$y^2 - 7xy - 36y = x^3 - 18x^2$	$\mathbb{Z}/10$
$y^2 + 43xy - 210y = x^3 - 210x^2$	$\mathbb{Z}/12$
$y^2 = x^3 - x$	$\mathbb{Z}/2 \oplus \mathbb{Z}/2$
$y^2 = x^3 + 5x^2 + 4x$	$\mathbb{Z}/2 \oplus \mathbb{Z}/4$
$y^2 + 5xy - 6y = x^3 - 3x^2$	$\mathbb{Z}/2 \oplus \mathbb{Z}/6$
$y^2 = x^3 + 337x^2 + 20736x$	$\mathbb{Z}/2 \oplus \mathbb{Z}/8$

§ 6. Гипотеза Таниямы и последняя теорема Ферма

Во введении к настоящей главе мы процитировали замечание Пьера Ферма о невозможности представить целое число, являющееся степенью большей двух, в виде суммы двух таких же степеней. Этот текст на латыни, следующий за указанием «Наблюдение господина де Ферма», находится в издании трудов Ди-

офанта, которое было выпущено Ферма-сыном в 1670 году через пять лет после смерти его отца. Оно доподлинно воспроизводит замечание, внесенное Ферма в его собственный экземпляр сочинений Диофанта, в настоящее время утраченный.

Ферма утверждал, что уравнение

$$a^n + b^n = c^n$$

не имеет целочисленных решений с $abc \neq 0$. Это утверждение называется *последней* (или *великой*) *теоремой Ферма*. За триста лет оно было проверено для всех $n \leqslant 150000$, но до последнего времени в общем случае оставалось не доказанным.

В этом параграфе мы опишем подход к доказательству теоремы Ферма, основанный на теории эллиптических кривых.

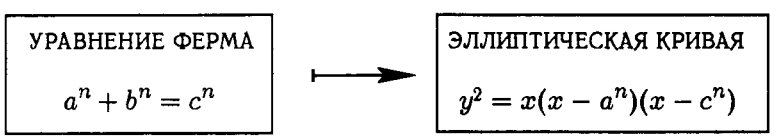
История взаимосвязи между последней теоремы Ферма и эллиптическими кривыми начинается в 1955 году, когда японский математик Ютака Танияма (1927–1958) сформулировал проблему, представляющую собой несколько ослабленную версию следующей гипотезы.

Гипотеза 1 (Танияма). *Любая эллиптическая кривая, определенная над полем рациональных чисел, является модулярной.*

В такой форме гипотеза Таниямы появилась в начале шестидесятых годов в работах японца Горо Шимуры. В последующие годы Шимурой и французским математиком Андрэ Вейлем была показана фундаментальная связь гипотезы Таниямы со многими разделами арифметики эллиптических кривых. На рубеже шестидесятых и семидесятых годов французский математик Ив Эллегарш сопоставил с уравнением Ферма $a^n + b^n = c^n$ эллиптическую кривую

$$y^2 = x(x - a^n)(x - c^n) \tag{6.1}$$

и использовал известные к тому времени результаты, касающиеся теоремы Ферма, для изучения точек конечного порядка на эллиптических кривых. Дальнейшее развитие событий показало, что сопоставление



является поистине революционным. В 1985 г. немецкий математик Герхард Фрей предположил, что эллиптическая кривая, соответствующая контрпримеру к теореме Ферма, не может быть модулярной (в противоречие с гипотезой Таниямы). Самому Фрею не удалось доказать это утверждение, однако вскоре доказательство было получено американским математиком Кеннетом Рибетом. Другими словами, Рибет показал, что *последняя теорема Ферма является следствием гипотезы Таниямы.*

23 июня 1993 г. математик из Принстона Эндрю Уайлс, выступая на конференции по теории чисел в Кембридже (Великобритания), анонсировал доказательство гипотезы Таниямы для широкого класса эллиптических кривых (так называемых полустабильных кривых), в который, в частности, входят все кривые вида (6.1). Тем самым он заявил, что доказал последнюю теорему Ферма. Дальнейшие события развивались довольно драматически. В начале декабря 1993 г. за несколько дней до того, как рукопись работы Уайлса должна была пойти в печать, в его доказательстве были обнаружены пробелы. Исправление их заняло свыше года. Текст с доказательством гипотезы Таниямы, написанный Уайлсом в сотрудничестве с Тейлором, вышел в свет летом 1995 года [B17, B19].

В рамках нашей небольшой книги мы не имеем возможности сколь-нибудь подробно обсудить гипотезу Таниямы и привести ее доказательство. Поэтому мы ограничимся тем, что сформулируем все основные понятия, относящиеся к гипотезе Таниямы, и покажем, как из этой гипотезы вытекает последняя теорема Ферма.

Начнем с понятия *модулярности*. Пусть H — верхняя комплексная полуплоскость

$$H = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}.$$

Для натурального числа N определим группу матриц $\Gamma_0(N)$, полагая

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}), \mid c \equiv 0 \pmod N \right\}.$$

Группа $\Gamma_0(N)$ действует на H с помощью дробно-линейных преобразований

$$g\tau = \frac{a\tau + b}{c\tau + d}, \qquad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N), \quad \tau \in H.$$

Нетрудно показать, что факторпространство $H/\Gamma_0(N)$ является некомпактной римановой поверхностью. Мы хотели бы пополнить эту поверхность, т. е. превратить ее в компактную. Напомним, прежде всего, что группа, действующая на множестве, разбивает это множество на классы эквивалентности: две точки лежат в одном классе, если существует элемент из группы, переводящий одну из этих точек в другую. В частности, если G — подгруппа в $SL_2(\mathbb{Z})$, то мы говорим, что две точки $\tau_1, \tau_2 \in H$ являются G-эквивалентными, если существует такой элемент $g \in G$, что $\tau_2 = g\tau_1$.

Обозначим через $\overline{H}$ множество $H \cup \{\infty\} \cup \mathbb{Q}$, где H — верхняя полуплоскость. Это означает, что мы добавляем к H бесконечную точку (которую можно представить себе как точку, находящуюся очень высоко на мнимой оси), а также все рациональные точки на вещественной оси. Точки $\{\infty\} \cup \mathbb{Q}$ называются *параболическими*. Легко видеть, что группа $SL_2(\mathbb{Z})$ транзитивно действует на параболических точках. В самом деле, запишем некоторое рациональное число в виде дроби a/c со взаимно простыми a и c. Мы можем найти b и d из уравнения $ad - bc = 1$ и построить матрицу $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$. Эта матрица переводит ∞ в a/c. Следовательно, все рациональные числа попадают в тот же класс $SL_2(\mathbb{Z})$-эквивалентности, что и ∞. Подгруппа $\Gamma_0(N)$ также переставляет параболические точки, но уже не транзитивно. Это означает, что множество параболических точек $\{\infty\}$ и $\mathbb{Q}$ разбивается больше чем на один класс эквивалентности. Например, если p — простое число, то $\Gamma_0(p)$ имеет две параболические точки ∞ и 0, в то же время $\Gamma_0(p^2)$ имеет $p+1$ параболическую точку ∞, 0 и $-\dfrac{1}{kp}$, где $k = 1, \ldots, p-1$.

Обычная топология на H продолжается на множество $\overline{H}$ следующим образом. Фундаментальную систему окрестностей точки ∞ составляют множества $U_C = \{\tau \in H \mid \operatorname{Im}\tau > C\} \cup \{\infty\}$ для всех $C > 0$.

Рассмотрим теперь параболическую точку $a/c \in \mathbb{Q} \subset \overline{H}$. Дополним пару чисел a, c до матрицы

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

С помощью отображения g переведем окрестность U_C в круги, касающиеся вещественной оси в точке a/c. Эти круги образуют фундаментальную систему открытых окрестностей точки a/c. Другими словами, последовательность τ_j сходится в этой топологии к a/c тогда и только тогда, когда последовательность чисел $\operatorname{Im} g^{-1}(\tau_j)$ сходится в обычной топологии к бесконечности.

Положим $X_0(N) = \overline{H}/\Gamma_0(N)$. можно показать, что $X_0(N)$ — компактная риманова поверхность. Ее род задается формулой

$$g = g(X_0(N)) = 1 + \frac{1}{2}\mu(N) - \frac{1}{4}\mu_2(N) - \frac{1}{3}\mu_3(N) - \frac{1}{2}\mu_\infty(N), \quad (6.2)$$

где

$$\mu(N) = \begin{cases} 3, & \text{если } N = 2, \\ N \prod\limits_{p \mid N} \left(1 + \dfrac{1}{p}\right), & \text{если } N > 2; \end{cases}$$

$$\mu_2(N) = \begin{cases} 0, & \text{если } N \equiv 0 \pmod 4, \\ N \prod\limits_{p \mid N} \left(1 + \left(\dfrac{-1}{p}\right)\right), & \text{в противном случае;} \end{cases}$$

$$\mu_3(N) = \begin{cases} 0, & \text{если } N \equiv 0 \pmod 2 \text{ или } N \equiv 0 \pmod 9, \\ N \prod\limits_{p \mid N} \left(1 + \left(\dfrac{-3}{p}\right)\right), & \text{в противном случае;} \end{cases}$$

$$\mu_\infty(N) = \sum_{d \mid N} \varphi\left(\left(d, \frac{N}{d}\right)\right).$$

В этих формулах мы подразумеваем, что $\varphi(1) = 1$, и $\left(\dfrac{a}{p}\right)$ — символ Лежандра, т. е.

$$\left(\frac{-1}{p}\right) = \begin{cases} 0, & \text{если } p = 2, \\ 1, & \text{если } p \equiv 1 \pmod 4, \\ -1, & \text{если } p \equiv 3 \pmod 4; \end{cases}$$

$$\left(\frac{-3}{p}\right) = \begin{cases} 0, & \text{если } p = 3, \\ 1, & \text{если } p \equiv 1 \pmod 3, \\ -1, & \text{если } p \equiv 2 \pmod 3. \end{cases}$$

При малых N род $g = g(X_0(N))$ представлен в следующей таблице:

Табл. 1

N	1 ... 10	11	12	13	14	15	16	17	18	19
g	0	1	0	0	1	1	0	1	0	1
N	20	21	22	23	24	25	26	27	28	29
g	1	1	2	2	1	0	2	1	2	2

Выше мы отмечали, что любая эллиптическая кривая является римановой поверхностью рода 1.

Назовем эллиптическую кривую E *модулярной*, если для некоторого натурального числа N существует голоморфное отображение из $X_0(N)$ на E.

Еще одно понятие, с которым мы встречаемся при выводе последней теоремы Ферма из гипотезы Таниямы, — это *полустабильные кривые.*

Пусть E — некоторая эллиптическая кривая, заданная уравнением

$$y^2 = x^3 + ax + b, \tag{6.3}$$

в котором a и b — целые числа. Как и в предыдущем параграфе, для простого числа p обозначим через E_p редукцию кривой E по модулю p, т. е. кривую

$$y^2 = x^3 + \bar{a}x + \bar{b}, \qquad \bar{a},\ \bar{b} \in \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p,$$

Рассмотрим те простые числа p, которые делят дискриминант $\Delta = -(4a^3+27b^2)$ кривой (т. е. те p, для которых редукция E_p не является эллиптической кривой над соответствующим полем $\mathbb{F}_p$). Для таких p многочлен $x^3 + \bar{a}x + \bar{b}$ имеет кратный корень и, следовательно, кривая E_p содержит особую точку. Если кратность корня равна двум, то через эту точку проходят две различные касательные к E_p, если же кратность равна трем, то имеются две совпадающие касательные. В первом случае особая точка называется *точкой самопересечения*, а во втором — *точкой возврата.*

Стоит отметить, что точно так же можно классифицировать особые точки редукций и для эллиптических кривых, заданных уравнениями вида

$$y^2 + a_1xy + a_3 = x^3 + a_2x^2 + a_4x + a_6.$$

с целыми коэффициентами, поскольку подходящими заменами переменных

$$x = u^2x' + r, \quad y = u^3y' + su^2x' + t, \quad u, r, s, t \in \mathbb{Q},$$

эти уравнения преобразуются к уравнениям вида (6.3) (см. § 4 гл. 1).

Эллиптическая кривая E над полем $\mathbb{Q}$ называется *полустабильной*, если для всех простых p ее редукции E_p не имеют особых точек возврата.

Теперь мы в состоянии точно сформулировать утверждение, доказанное Эндрю Уайлсом.

Гипотеза 2 (Гипотеза Таниямы для полустабильных кривых). *Всякая полустабильная эллиптическая кривая, определенная над полем рациональных чисел, является модулярной.*

Вернемся вновь к определению модулярности. Можно доказать, что фигурирующее в нем натуральное число N — это так называемый *кондуктор эллиптической кривой.* Он является делителем ее дискриминанта и для произвольной кривой определяется довольно сложно. В случае же полустабильных эллиптических кривых нетрудно получить его полное описание. Начнем с нескольких вспомогательных замечаний.

Пусть r — некоторое рациональное и p — простое число. Если $r \neq 0$, представим его в виде $r = p^n\dfrac{u}{v}$, где $u, v \in \mathbb{Z}$ и $(u,p) = (v,p) = 1$. Положим

$$|r|_p = p^{-n}, \quad r \neq 0; \qquad |0|_p = 1.$$

Число $|r|_p$ называется *p-адической нормой* числа r и обладает следующими свойствами:

1) $|r+s|_p \leqslant \max\{|r|_p,\ |s|_p\}$,

2) $|rs|_p = |r|_p\,|s|_p$.

Скажем, что число $r \in \mathbb{Q}$ является *p-целым*, если $|r|_p \leqslant 1$.

Рассмотрим теперь некоторую эллиптическую кривую. Уравнение такой кривой можно записать в виде

$$y^2 + a_1 y + a_3 y = x^3 + a_2 x + a_4 x + a_6, \tag{6.4}$$

откуда линейными заменами получается форма Вейерштрасса. Прямые вычисления показывают, что дискриминант кривой (6.4) равен

$$\Delta = -b_2^2 b_8 - 8b_4^3 - 27b_6^2 + 9b_2 b_4 b_6, \tag{6.5}$$

где

$$b_2 = a_1^2 + 4a_2, \quad b_4 = 2a_4 + a_1 a_3, \quad b_6 = a_3^2 + 4a_6,$$
$$b_8 = a_1^2 a_6 + 4a_2 a_6 - a_1 a_3 a_4 + a_2 a_3^2 - a_4^2.$$

В частности, замена переменных

$$(x, y) \mapsto \left(\frac{x - 3b_2}{36}, \frac{y}{216}\right)$$

приводит уравнение (6.4) к виду

$$y^2 = x^3 - 27c_4 x - 54c_6, \tag{6.6}$$

где $c_4 = b_2^2 - 27b_4$, $c_6 = -b_2^3 + 36b_2 b_4 - 216b_6$. Для дискриминанта кривой (6.6) имеет место соотношение

$$1728\,\Delta = c_4^3 - c_6^2. \tag{6.7}$$

Назовем уравнение (6.4) с целыми коэффициентами *минимальным для простого числа* p, если p-адическая норма $|\Delta|_p$ дискриминанта (6.5) не возрастает при заменах переменных

$$x = u^2 x' + r, \quad y = u^3 y' + s u^2 x' + t, \qquad u, r, s, t \in \mathbb{Q}, \tag{6.8}$$

обладающих тем свойством, что новые коэффициенты в уравнении кривой являются p-целыми. Уравнение (6.4) называется уравнением эллиптической кривой в *глобально минимальной форме*, если оно минимально для всех простых и его коэффициенты — целые числа.

Можно показать, что для любой эллиптической кривой над существует такая замена переменных (6.8), что результирующее уравнение является глобально минимальным. Мы не станем доказывать это утверждение. Вместо этого мы покажем, как можно

конструктивно достичь минимальности одновременно для всех простых $p > 3$.

Лемма 1. *Пусть p — простое число и все коэффициенты a_i в уравнении* (6.4) *являются p-целыми. Если выполняется одно из трех неравенств*

$$|\Delta|_p > p^{-12}, \quad |c_4|_p > p^{-4}, \quad |c_6|_p > p^{-6},$$

то уравнение (6.4) *минимально для простого p. Если же $p > 3$, $|\Delta|_p \leqslant p^{-12}$ и $|c_4|_p \leqslant p^{-4}$, то это уравнение не минимально для p.*

Доказательство. Предположим, что замена переменных (6.8) приводит к такой системе p-целых коэффициентов $\{a_i'\}$, что $1 \geqslant |\Delta'|_p > |\Delta|_p$. Прямое вычисление показывает, что $u^{12}\Delta' = \Delta$, поэтому $|u|_p^{12}\,|\Delta'|_p = |\Delta|_p$, откуда $|u|_p < 1$. Но тогда $|u|_p \leqslant p^{-1}$ и

$$|\Delta|_p = |u|_p^{12}\,|\Delta'|_p \leqslant p^{-12}.$$

Рассуждения для c_4 и c_6 аналогичны. Обратно, пусть $p > 3$, $|\Delta|_p \leqslant p^{-12}$ и $|c_4|_p \leqslant p^{-4}$. Согласно (6.7)

$$1728\,\Delta = c_4^3 - c_6^2.$$

Так как $|1728|_p = 1$, то $|c_6|_p \leqslant p^{-6}$. Ранее мы отмечали, что существует замена переменных, переводящая уравнение (6.4) в уравнение (6.6):

$$y^2 = x^3 - 27c_4 x - 54c_6$$

с дискриминантом $\Delta' = 2^{12}3^{12}\Delta$. Если мы применим к (6.6) замену переменных вида (6.8) с $u = p$ и $r = s = t = 0$, то в результате получим уравнение

$$y^2 = x^3 - 27xc_4 p^{-4} - 54c_6 p^{-6}.$$

Поскольку $|c_4 p^{-4}|_p \leqslant 1$, $|c_6 p^{-6}|_p \leqslant 1$, это уравнение имеет p-целые коэффициенты. Его дискриминант Δ'' имеет p-адическую норму $|\Delta''|_p = p^{12}|\Delta'|_p = p^{12}|\Delta|_p$. Следовательно, данное уравнение не было минимальным для p, что и требовалось доказать. □

Пусть Δ — дискриминант полустабильной эллиптической кривой E, заданной в глобально минимальной форме, и

$$\Delta = p_1^{\delta_1} \ldots p_k^{\delta_k}$$

— разложение модуля Δ на простые множители. Тогда *кондуктор N кривой E* определяется формулой

$$N = p_1 \dots p_k. \tag{6.9}$$

Для того, чтобы «перебросить мостик» между гипотезой Таниямы и последней теоремой Ферма, нам необходимо переформулировать «бескоординатное» геометрическое определение модулярности в теоретико-числовых терминах. Это достигается с помощью одного из наиболее могущественных инструментов современной теории чисел: модулярных форм.

Как обычно, через H будем обозначать верхнюю комплексную полуплоскость.

Пусть N — натуральное число и k — целое число. *Модулярной формой* веса k для группы $\Gamma_0(N)$ называется голоморфная функция $f\colon H \to \mathbb{C}$, удовлетворяющая условию

$$f(g\tau) = (c\tau + d)^k f(\tau) \tag{6.10}$$

для любых $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ и $\tau \in H$, и голоморфная во всех параболических точках.

Подставив в уравнение (6.5) матрицу $g = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N)$, мы получим, что $f(\tau) = f(\tau + 1)$ для всех $\tau \in H$. Следовательно, функция f допускает разложение в ряд Фурье

$$f(\tau) = \sum_{n=0}^{\infty} a_n q^n, \quad q = e^{2\pi i \tau}, \tag{6.11}$$

с комплексными коэффициентами a_n, причем все a_n с отрицательными номерами n равны нулю, поскольку f голоморфна в параболической точке $\{\infty\}$. Назовем модулярную форму f *параболической*, если f обращается в нуль во всех параболических точках. В частности, для параболической формы коэффициент a_0 (значение в точке $\{\infty\}$) в разложении (6.11) равен нулю. Назовем параболическую форму нормализованной, если $a_1 = 1$.

Пространство параболических форм веса k для группы $\Gamma_0(N)$ обозначается через $S_k(\Gamma_0(N))$. Использование буквы S традиционно: это первая буква немецкого слова spitzenform — параболическая форма.

Очевидно, что $\dim(S_k(\Gamma_0(N))) = 0$ для нечетных k. Можно доказать, что все пространства $S_k(\Gamma_0(N))$ конечномерны.

В дальнейшем нас будут особо интересовать параболические формы веса 2 для группы $\Gamma_0(N)$. Каждая такая форма определяет голоморфный дифференциал на римановой поверхности $X_0(N) = \overline{H}/\Gamma_0(N)$, поскольку

$$d\left(\frac{a\tau + b}{c\tau + d}\right) = \frac{d\tau}{(c\tau + d)^2}$$

и, следовательно, $f(\tau)d\tau$ инвариантен относительно действия группы $\Gamma_0(N)$ на $\overline{H}$. Так как пространство параболических форм $S_2(\Gamma_0(N))$ совпадает с пространством голоморфных дифференциалов на $X_0(N)$, то

$$\dim(S_2(\Gamma_0(N))) = g(X_0(N)) = g, \tag{6.12}$$

где род g задается формулой (6.2). В частности (см. табл. 1),

$$\dim(S_2(\Gamma_0(2))) = 0. \tag{6.12}$$

Эта нехитрая формула и сыграет важную роль при выводе из гипотезы Таниямы последней теоремы Ферма.

Для фиксированного N на пространстве $S_2(\Gamma_0(N))$ имеется семейство линейных операторов

$$T_m\colon\ S_2(\Gamma_0(N)) \to S_2(\Gamma_0(N)),$$

индексированных целыми числами $m \geqslant 1$. Эти операторы задаются следующим образом. Пусть

$$f(\tau) = \sum_{n=1}^{\infty} a_n q^n$$

— ряд Фурье параболической формы $f(\tau) \in S_2(\Gamma_0(N))$. Тогда

$$T_m f(\tau) = \sum_{n=1}^{\infty} \Bigg(\sum_{\substack{(d,N)=1 \\ d \mid (n,m)}} d a_{nm/d^2} \Bigg) q^n. \tag{6.14}$$

Операторы T_m называются *операторами Гекке*.

Собственной формой называется нормализованная параболическая форма $f(\tau) \in S_2(\Gamma_0(N))$, являющаяся собственной функцией для всех операторов Гекке T_m.

Из формулы (6.9) следует, что если

$$f(\tau) = \sum_{n=1}^{\infty} a_n q^n$$

— собственная форма, то $T_m f = a_m f$ для всех $m \geqslant 1$.

Если

$$f(\tau) = \sum_{n=1}^{\infty} a_n q^n$$

— параболическая форма из $S_k(\Gamma_0(N))$, то с ней можно связать L-функцию $L(f,s)$, задаваемую при $\operatorname{Re} s > k/2 + 1$ рядом

$$L(f,s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s} \tag{6.15}$$

и аналитически продолженную на всю комплексную плоскость.

Используя явный вид операторов Гекке (6.14), мы можем легко показать, что нормализованная собственная параболическая форма $f(\tau) \in S_2(\Gamma_0(N))$ допускает следующее разложение в бесконечное произведение:

$$L(f,s) = \prod_{p \mid N} \left(\frac{1}{1 - a_p p^{-s}} \right) \cdot \prod_{p \nmid N} \left(\frac{1}{1 - a_p p^{-3} + p^{1-2s}} \right) \tag{6.16}$$

Отметим полное формальное сходство этого разложения с разложением (5.2) для L-функции $L(E,s)$ эллиптической кривой E.

В терминах L-функции $L(f,s)$ определение модулярности формулируется следующим образом. Эллиптическая кривая E над полем рациональных чисел называется *модулярной*, если существует собственная параболическая форма $f(\tau) = \sum_{n=1}^{\infty} a_n q^n \in S_2(\Gamma_0(N))$, где N — кондуктор кривой, такая что $L(f,s) = L(E,s)$.

Другими словами, эллиптическая кривая E модулярна, если существует собственная модулярная форма $f(\tau) = \sum_{n=1}^{\infty} a_n q^n \in S_2(\Gamma_0(N))$, такая что

$$a_n = p + 1 - N_p, \tag{6.17}$$

где N_p — число точек редукции E_p, включая бесконечно удаленную точку.

Теперь мы готовы к штурму теоремы Ферма. Очевидно, что если эта теорема доказана для некоторого показателя n, то тем самым она доказана и для всех показателей, кратных n. Так как всякое целое число $n > 2$ делится или на 4, или на нечетное простое число, то можно поэтому ограничиться случаем, когда показатель равен либо четырем, либо нечетному простому числу. Для $n = 4$ элементарное доказательство теоремы Ферма было получено Эйлером. Таким образом, достаточно изучить уравнение

$$a^l + b^l = c^l \tag{6.18}$$

в котором показатель l есть нечетное простое число. Очевидно, что в уравнении (6.18) числа a, b и c можно считать попарно взаимно простыми.

Пусть

$$a^l + b^l = c^l$$

— контрпример к теореме Ферма, и пусть $l \geqslant 5$. Рассмотрим эллиптическую кривую E, заданную уравнением

$$y^2 = x(x - a^l)(x - c^l).$$

Дискриминант этой кривой равен

$$\Delta = 16\, a^{2l} b^{2l} c^{2l},$$

а в представлении (6.6) имеем

$$c_4 = 16(a^{2l} - a^l c^l + c^{2l}).$$

Представим кривую E в глобально минимальной форме и вычислим ее кондуктор N. Любое нечетное простое число p, делящее дискриминант Δ, делит произведение abc. Так как числа a, b и c попарно взаимно просты, то отсюда следует, что коэффициент c_4 не делится на p. Поэтому в силу леммы 1 эллиптическая кривая минимальна для p. Если $p \mid ac$, то прямая проверка показывает, что (0,0) является точкой самопересечения для кривой E_p. Если же $p \mid b$, то E_p имеет точку самопересечения $(a^l, 0)$. Таким образом, кондуктор N должен иметь вид

$$N = 2^{\alpha} \prod_{\substack{p \mid abc \\ p \text{ нечетно}}} p, \tag{6.19}$$

и всё что остается, это вычислить показатель α (т. е. выяснить, равен он нулю или единице). Без ограничения общности можно считать, что число c четно, поэтому

$$a^l \equiv 0 \pmod{32}. \tag{6.20}$$

Так как $b^l \equiv -a^l \pmod 4$ и l нечетно, то, меняя в случае необходимости местами a и b, мы можем считать, что

$$a^l \equiv 1 \pmod 4. \tag{6.21}$$

Полагая $x = 4X$ и $y = 8Y + 4X$, мы приходим к уравнению

$$Y^2 + XY = X^3 + \frac{1}{4}(1 - a^l - c^l)X^2 + \frac{1}{16}a^l c^l X. \tag{6.22}$$

Сравнения (6.20) и (6.21) показывают, что коэффициенты этого уравнения целые, и поэтому оно задает глобально минимальную форму нашей эллиптической кривой. По модулю 2 уравнение (6.22) имеет вид

$$Y^2 + XY = X^3$$

или

$$Y^2 + XY = X^3 + X^2.$$

с особой точкой $(X, Y) = (0, 0)$. Так как ни один из многочленов $Y^2 + XY$, $Y^2 + XY + Y^2$ не является полным квадратом, точка $(0, 0)$ есть точка самопересечения.

Принимая во внимание соотношение (6.19), мы получаем, что кондуктор нашей кривой задается формулой

$$N = \prod_{p \mid abc} p. \tag{6.23}$$

Воспользуемся теперь следующей теоремой.

Теорема 2 (Рибет). *Пусть E — эллиптическая кривая над полем $\mathbb{Q}$, заданная в глобально минимальной форме и имеющая дискриминант*

$$\Delta = \prod_{p \mid \Delta} p^{\delta_p}$$

и кондуктор

$$N = \prod_{p \mid \Delta} p^{\varepsilon_p}.$$

Предположим, что E — модулярная кривая с модулярной параметризацией уровня N, заданной нормализованной параболической формой

$$f(\tau) = q + \sum_{n=2}^{\infty} a_n q^n \in S_2(\Gamma_0(N)).$$

Фиксируем простое число l, и пусть

$$N_1 = \frac{N}{\left(\prod\limits_{p:\ \varepsilon_p = 1;\ l \mid \delta_p} p\right)}. \tag{6.24}$$

Тогда существует параболическая форма

$$f_1(\tau) = \sum_{n=1}^{\infty} \alpha_n q^n \in S_2(\Gamma_0(N_1))$$

с целыми коэффициентами, такая что

$$a_n \equiv \alpha_n \pmod l$$

для всех $1 \leqslant n < \infty$.

Теорема Рибета сводит нахождение связей между гипотезой Таниямы и теоремой Ферма к весьма простым вычислениям.

Теорема 3. *Из гипотезы Таниямы для полустабильных эллиптических кривых следует последняя теорема Ферма.*

Доказательство. Предположим, что теорема Ферма неверна, и пусть

$$a^l + b^l = c^l$$

— соответствующий контрпример (как и выше, здесь l — нечетное простое число, большее трех, и a, b, c — попарно взаимно просты). Применим теорему 2 к эллиптической кривой

$$y^2 = x(x - a^l)(x - c^l).$$

Сравнивая формулы (6.23) и (6.24), мы видим, что $N_1 = 2$. Следовательно, по теореме 2 найдется параболическая форма

$$f_1(\tau) = \sum_{n=1}^{\infty} d_n q^n,$$

лежащая в $S_2(\Gamma_0(2))$. Но в силу (6.8) — это нулевое пространство. Поэтому $d_n = 0$ для всех n. В то же время $a_1 = 1$. Стало быть, сравнение

$$a_n \equiv d_n \pmod{l}$$

не выполняется для $n = 1$, и мы приходим к противоречию. □

Заметим, в заключение, что значение гипотезы Таниямы отнюдь не ограничивается ее связью с теоремой Ферма. Как уже кратко упоминалось в предыдущем параграфе, из гипотезы Таниямы вытекает гипотеза Хассе — Вейля (для эллиптических кривых над полем рациональных чисел эти гипотезы эквивалентны), а вместе с нею открываются новые горизонты в исследовании арифметики эллиптических кривых.

ГЛАВА 6

АЛГЕБРАИЧЕСКИЕ УРАВНЕНИЯ

В последних двух главах мы займемся алгебраическими уравнениями. Проблема решения уравнений в радикалах интересовала математиков всех времен. После того, как Тарталья и Феррари решили уравнения третьей и четвертой степени, появились надежды решить любое алгебраическое уравнение, да и сама алгебра до конца XVIII столетия развивалась как наука о решении уравнений. Эти надежды серьезно поколебали Жозеф Луи Лагранж и Паоло Руффини, а окончательно их развеял Нильс Хенрик Абель [A1], доказав в 1824 г. невозможность решения общего уравнения пятой степени в радикалах.

Дальнейшие продвижения теория алгебраических уравнений получила в направлении трансцендентного анализа. Творцом этого направления, по-видимому, можно считать Франсуа Виета, который еще в конце XVI в., изучая проблему трисекции угла, предложил аналитический способ решения кубического уравнения

$$x^3 - 3x + a = 0,$$

где a — положительное число, не превосходящее двух. Виет, рассматривая для этого уравнения подстановку $a = 2\sin\alpha$, показал, что уравнению удовлетворяет величина $x = 2\sin(\alpha/3)$. Идея Виета ждала своего развития примерно 275 лет. В 1858 г. молодой французский математик Шарль Эрмит [A14, a], занимаясь теорией модулярных уравнений, пришел к выводу, что всякое уравнение пятой степени можно решить в модулярных эллиптических функциях. Одновременно к аналогичным результатам пришел Леопольд Кронекер, преобразовав уравнение пятой степени в модулярное уравнений шестой степени.

§ 1. Решение уравнений 3-й и 4-й степени

Известно много разных способов решения в радикалах уравнений 3-й и 4-й степени. В этом параграфе мы обсудим лишь наиболее простые из них. Некоторые другие способы решения мы обсудим в § 3 и § 6.

Предварительно заметим, что уравнение

$$x^n + a_1 x^{n-1} + \ldots + a_n = 0$$

можно привести к виду

$$y^n + b_2 y^{n-2} + \ldots + b_n = 0$$

с помощью замены $y = x + \frac{a_1}{n}$. Поэтому достаточно рассмотреть кубические уравнения вида $x^3 + ax + b = 0$ и уравнения 4-й степени вида $x^4 + ax^2 + bx + c = 0$.

В первую очередь рассмотрим кубические уравнения. Будем искать корни уравнения

$$x^3 + ax + b = 0 \tag{1.1}$$

в виде $x = \sqrt[3]{p} + \sqrt[3]{q}$. Тогда

$$x^3 = p + q + 3\sqrt[3]{pq}\left(\sqrt[3]{p} + \sqrt[3]{q}\right) = p + q + 3\sqrt[3]{pq}\,x.$$

Поэтому p и q должны быть такими, чтобы выполнялись равенства $3\sqrt[3]{pq} = -a$ и $p + q = -b$. Тогда p и q являются корнями уравнения

$$t^2 + bt - \frac{a^3}{27} = 0.$$

Следовательно,

$$p, q = -\frac{b}{2} \pm \sqrt{\frac{b^2}{4} + \frac{a^3}{27}} \tag{1.2}$$

Формула $x = \sqrt[3]{p} + \sqrt[3]{q}$ дает 9 различных значений. Чтобы получить формулу, которая дает 3 значения, можно воспользоваться соотношением $\sqrt[3]{p}\sqrt[3]{q} = -a/3$. Эта формула имеет вид

$$x = \sqrt[3]{p} - \frac{a}{3\sqrt[3]{p}}, \tag{1.3}$$

где p находится по формуле (1.2). При этом значение x не зависит от выбора знака перед радикалом в формуле (1.2).

Легко проверить, что определенные по формуле (1.2) значения x являются корнями уравнения (1.1).

Обратимся теперь к уравнениям 4-й степени.

Первый способ. Попытаемся представить многочлен $x^4 + ax^2 + bx + c$ в виде разности двух квадратов. Воспользуемся для этого тем, что

$$x^4 + ax^2 + bx + c = \left(x^2 + \frac{a}{2} + t\right)^2 - \left(2tx^2 - bx + \left(t^2 + at - c + \frac{a^2}{2}\right)\right).$$

Выберем величину t так, чтобы дискриминант

$$D = b^2 - 8t\left(t^2 + at - c + \frac{a^2}{4}\right)$$

был равен нулю. Тогда

$$x^4 + ax^2 + bx + c = \left(x^2 + \frac{a}{2} + t\right)^2 - 2t\left(x - \frac{b}{4t}\right)^2.$$

Поэтому уравнение

$$x^4 + ax^2 + bx + c = 0 \tag{1.4}$$

можно решить следующим образом. Сначала решим относительно t кубическое уравнение

$$b^2 - 8t\left(t^2 + at - c + \frac{a^2}{4}\right) = 0.$$

Пусть t_0 — один из его корней. Тогда уравнение (1.4) можно записать в виде

$$x^2 + \frac{a}{2} + t_0 = \pm\sqrt{2t_0}\left(x - \frac{b}{4t_0}\right).$$

Второй способ принадлежит Эйлеру. Пусть x_1, x_2, x_3, x_4 — корни уравнения (1.4). Положим $u = x_1 + x_2 = -(x_3 + x_4)$. Тогда

$$x^4 + ax^2 + bx + c = (x^2 - ux + \alpha)(x^2 + ux + \beta),$$

т. е.

$$\alpha + \beta - u^2 = a, \qquad u(\alpha - \beta) = b, \qquad \alpha\beta = c.$$

Из первого и второго уравнений получаем

$$\alpha = \frac{1}{2}\left(a + u^2 + \frac{b}{u}\right), \qquad \beta = \frac{1}{2}\left(a + u^2 - \frac{b}{u}\right).$$

Подставив эти выражения в третье уравнение, получим

$$u^6 + 2au^4 + (a^2 - 4c)\,u^2 - b^2 = 0. \tag{1.5}$$

Уравнение (1.5) является кубическим уравнением относительно u^2. Решив это кубическое уравнение, найдем 6 корней уравнения (1.5). Они имеют вид $\pm u_1$, $\pm u_2$, $\pm u_3$. Можно считать, что

$$\begin{aligned} x_1 + x_2 &= u_1, & x_3 + x_4 &= -u_1, \\ x_1 + x_3 &= u_2, & x_2 + x_4 &= -u_2, \\ x_1 + x_4 &= u_3, & x_2 + x_3 &= -u_3. \end{aligned}$$

Тогда $u_1 + u_2 + u_3 = 2x_1$.

Третий способ. Рассмотрим систему уравнений

$$\begin{cases} f = y - x^2 = 0 \\ g = y^2 + ay + bx + c = 0 \end{cases}$$

Второе уравнение можно заменить на $\lambda f + g = 0$. Подберем λ так, чтобы $\lambda f + g$ оказалось произведением двух линейных функций. Это означает, что кривая $\lambda f + g = 0$ представляет собой пару прямых. Так бывает тогда и только тогда, когда

$$\begin{vmatrix} -\lambda & 0 & \dfrac{b}{2} \\ 0 & 1 & \dfrac{a+\lambda}{2} \\ \dfrac{b}{2} & \dfrac{a+\lambda}{2} & c \end{vmatrix} = 0 \tag{1.6}$$

Поэтому если λ_0 — корень кубического уравнения (1.6), то уравнение $\lambda_0 f + g = 0$ распадается на два линейных уравнения, и система уравнений $f = 0$, $\lambda_0 f + g = 0$ легко решается.

§ 2. Симметрические многочлены

Многочлен $f(x_1, \ldots, x_n)$ называется *симметрическим*, если для любой перестановки $\begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix}$ выполняется равенство

$$f(x_{i_1}, \ldots, x_{i_n}) = f(x_1, \ldots, x_n)$$

Важнейшим примером симметрических многочленов служат *элементарные симметрические многочлены* $\sigma_i(x_1, \ldots, x_n)$, которые определяются соотношением

$$(x - x_1)\ldots(x - x_n) = x^n - \sigma_1 x^{n-1} + \sigma_2 x^{n-2} - \ldots + (-1)^n \sigma_n.$$

Таким образом, если $x_1, \ldots, x_n$ — корни многочлена $x^n + a_1 x^{n-1} + \ldots + a_n$, то

$$\sigma_i(x_1, \ldots, x_n) = (-1)^i a_i.$$

Удобно считать, что $\sigma_k(x_1, \ldots, x_n) = 0$ при $k > n$.

Теорема (Основная теорема о симметрических многочленах). *Любой симметрический многочлен $f(x_1, \ldots, x_n)$ можно представить в виде многочлена g от элементарных симметрических многочленов, т. е.*

$$f(x_1, \ldots, x_n) = g(\sigma_1, \ldots, \sigma_n).$$

При этом многочлен g определен однозначно.

Доказательство. Достаточно рассмотреть случай, когда многочлен f однородный. Будем говорить, что порядок члена $ax_1^{\alpha_1} \ldots x_n^{\alpha_n}$ выше порядка члена $bx_1^{\beta_1} \ldots x_n^{\beta_n}$, если $\alpha_1 = \beta_1, \ldots, \alpha_k = \beta_k$, $\alpha_{k+1} > \beta_{k+1}$ (возможно $k = 0$). Легко проверить следующие свойства этого порядка.

1) Наивысший член произведения двух многочленов равен произведению их наивысших членов.

2) Если $ax_1^{\alpha_1} \ldots x_n^{\alpha_n}$ — наивысший член симметрического многочлена, то $\alpha_1 \geqslant \alpha_2 \geqslant \ldots \geqslant \alpha_n$.

Пусть наивысший член однородного симметрического многочлена f равен $ax_1^{\alpha_1} \ldots x_n^{\alpha_n}$. Рассмотрим многочлен

$$a\sigma_1^{\alpha_1 - \alpha_2}\sigma_2^{\alpha_2 - \alpha_3} \ldots \sigma_n^{\alpha_n}.$$

Согласно свойству 1) его наивысший член равен

$$ax_1^{\alpha_1-\alpha_2}(x_1^{\alpha_2-\alpha_3}x_2^{\alpha_2-\alpha_3})\dots(x_1^{\alpha_n}\dots x_n^{\alpha_n}) = ax_1^{\alpha_1}\dots x_n^{\alpha_n}.$$

Потому порядок наивысшего члена у многочлена

$$f_1 = f - a\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3}\dots\sigma_n^{\alpha_n}$$

ниже, чем у многочлена f. Проделаем с многочленом f_1 такую же операцию, как с многочленом f, и т. д. Так как количество одночленов, порядок которых ниже порядка наивысшего члена многочлена f, конечно, то после нескольких операций получим нулевой многочлен. Это означает, что

$$f = a\sigma_1^{\alpha_1-\alpha_2}\sigma_2^{\alpha_2-\alpha_3}\dots\sigma_n^{\alpha_n} + \dots = g(\sigma_1\dots\sigma_n).$$

Докажем теперь единственность представления $f(x_1,\dots,x_n) = g(\sigma_1,\dots,\sigma_n)$. Для этого достаточно проверить, что если $g(\sigma_1,\dots,\sigma_n) = 0$ для любого набора $x_1,\dots,x_n$, то g — нулевой многочлен. Пусть g — ненулевой многочлен (над полем нулевой характеристики). Тогда $g(a_1,\dots,a_n) \neq 0$ для некоторого набора $a_1,\dots,a_n$. Ясно так же, что если $x_1,\dots,x_n$ — корни многочлена $x^n - a_1x^{n-1} + a_2x^{n-2} - \dots + (-1)^n a_n$, то $\sigma_i(x_1,\dots,x_n) = a_i$. □

Замечание. Если коэффициенты многочлена f целочисленные, то коэффициенты многочлена g тоже целочисленные.

Основная теорема о симметрических многочленах играет важную роль в теории алгебраических уравнений по следующей причине. Пусть $x_1,\dots,x_n$ — корни многочлена $x^n + a_1x^{n-1} + \dots + a_n$. Предположим, что некоторый многочлен $f(x_1,\dots,x_n)$ не изменяется при любых перестановках корней. Тогда его можно представить в виде многочлена от коэффициентов $a_1,\dots,a_n$. Этим свойством нам придется часто пользоваться.

Задачи.

1. Пусть $s_k(x_1,\dots,x_n) = x_1^k + \dots + x_n^k$. Доказать, что

$$s_k - s_{k-1}\sigma_1 + s_{k-2}\sigma_2 - \dots + (-1)^k k\sigma_k = 0$$

§ 3. Резольвенты Лагранжа

Квадратное уравнение $x^2 + ax + b = 0$ можно решать следующим образом. Пусть x_1 и x_2 — корни этого уравнения. Тогда

$$x_1 = \frac{1}{2}[(x_1+x_2)+(x_1-x_2)] = \frac{1}{2}\left[(x_1+x_2)+\sqrt{(x_1-x_2)^2}\right]. \quad (3.1)$$

При этом $x_1 + x_2$ и $(x_1 - x_2)^2$ — симметрические функции корней, а значит, их можно выразить через коэффициенты a и b. Выражения получатся следующие:

$$x_1 + x_2 = a,$$
$$(x_1 - x_2)^2 = (x_1 + x_2)^2 - 4x_1x_2 = a^2 - 4b.$$

Таким образом,

$$x_1 = \frac{1}{2}\left[-a + \sqrt{a^2 - 4b}\right].$$

Квадратный корень $\sqrt{a^2 - 4b}$ имеет два значения, дающих оба корня уравнения.

Аналогичный подход можно применить и к решению кубического уравнения. При этом вместо квадратных корней возникают кубические корни. Пусть α — примитивный кубический корень из единицы, т. е. $\alpha^3 = 1$ и $\alpha \neq 1$. Для корней кубического уравнения $x^3 + ax^3 + bx + c = 0$ формула, аналогичная формуле (3.1), выглядит следующим образом:

$$x_1 = \frac{1}{3}[(x_1+x_2+x_3)+(x_1+\alpha x_2+\alpha^2x_3)+(x_1+\alpha^2x_2+\alpha x_3)] =$$
$$= \frac{1}{3}(x_1+x_2+x_3)+$$
$$+\frac{1}{3}\left[\sqrt[3]{(x_1+\alpha x_2+\alpha^2x_3)^3}+\sqrt[3]{(x_1+\alpha^2x_2+\alpha x_3)^3}\right]. \quad (3.2)$$

Чтобы можно было воспользоваться этой формулой, нужно вычислить величины $u = (x_1 + \alpha x_2 + \alpha^2 x_3)^3$ и $v = (x_1 + \alpha^2 x_2 + \alpha x_3)^3$. При любой перестановке корней величина u переходит либо в себя, либо в v. В этом легко убедиться, если в полученном после перестановки корней выражении вынести за скобку

коэффициент α^k при x_1. Таким образом, $u+v$ и uv — симметрические многочлены от x_1, x_2 и x_3, а значит, их можно выразить через коэффициенты кубического уравнения. После того как эти выражения получены, решение кубического уравнения сводится к решению квадратного уравнения.

Такой метод решения кубического уравнения независимо предложили два французских математика, Лагранж и Вандермонд, причем сделали это они одновременно. В 1770 г. Вандермонд представил свою работу Парижской академии, а Лагранж, который с 1766 г. по 1787 г. жил и работал в Берлине, доложил свои результаты Берлинской академии. Работа Вандермонда «Mémoire sur la résolution des équations» содержала новый подход к решению уравнений третьей и четвертой степени, а также некоторых других уравнений, в том числе уравнения $x^{11}-1=0$. Статья Вандермонда была опубликована лишь в 1774 г. За это время, в 1771 г. и в 1773 г., были опубликованы две части фундаментального сочинения Лагранжа «Réflexions sur la résolution algébrique des équations» [A6]. В нем содержались почти те же идеи, что у Вандермонда, но они были разработаны существенно более детально. Вандермонд, ознакомившись с работами Лагранжа, к этой тематике больше не возвращался.

Основываясь на приведенном выше примере, введем следующее определение. Пусть $x^n + a_1x^{n-1} + \ldots + a_n$ — некоторый многочлен с рациональными коэффициентами, $x_0, \ldots, x_{n-1}$ — его корни. *Резольвентами Лагранжа* называются выражения

$$r(x_0, \alpha) = x_0 + \alpha x_1 + \ldots + \alpha^{n-1}x_{n-1},$$

где α — корень n-ой степени из единицы.

Точки α, удовлетворяющие условию $\alpha^n = 1$, расположены на комплексной плоскости в вершинах n-угольника, при этом точки α^k расположены в вершинах правильного m-угольника, где $m = n/(n,k)$. Поэтому

$$\sum_\alpha \alpha^k = \begin{cases} n & \text{при} \quad k = 0 \pmod n, \\ 0 & \text{при} \quad k \neq 0 \pmod n. \end{cases}$$

Следовательно,

$$\begin{aligned} nx_0 &= \sum_\alpha r(x_0, \alpha), \\ nx_k &= \sum_\alpha \alpha^{-k} r(x_0, \alpha). \end{aligned} \tag{3.3}$$

Таким образом, если резольвенты Лагранжа известны, то корни уравнения можно вычислить по формулам (3.3). Лагранж предложил следующий способ вычисления резольвент, который мы рассмотрим сначала для кубического уравнения. Пусть $r = x_0 + \alpha x_1 + \alpha^2 x_2$ — резольвента Лагранжа для кубического уравнения. В зависимости от порядка, в котором берутся корни x_0, x_1 и x_2, величина r принимает шесть значений $r_1, \ldots, r_6$. Рассмотрим многочлен шестой степени

$$g(t) = (t-r_1)(t-r_2)\ldots(t-r_6).$$

Его коэффициенты являются симметрическими многочленами от r_i, поэтому они являются симметрическими многочленами от x_0, x_1 и x_2, а значит, их можно выразить через коэффициенты исходного уравнения. Лагранж назвал уравнение $g(t)=0$ *разрешающим уравнением*. Дело в том, что оно легко решается в радикалах. В самом деле, пусть $r_1 = x_0 + \alpha x_1 + \alpha^2 x_2$ и $r_4 = x_0 + \alpha x_2 + \alpha^2 x_1$. Тогда остальные значения r_i равны $r_2 = \alpha r_1$, $r_3 = \alpha^2 r_1$, $r_5 = \alpha r_4$, $r_6 = \alpha^2 r_4$. Поэтому

$$(t-r_1)(t-r_2)(t-r_3) = t^3 - r_1^3,$$

$$(t-r_4)(t-r_5)(t-r_6) = t^3 - r_4^3,$$

а значит, $g(t) = (t^3 - r_1^3)(t^3 - r_4^3)$. Таким образом, уравнение $g(t)=0$ является квадратным уравнением относительно t^3.

Найдем разрешающее уравнение для многочлена четвертой степени. При этом в качестве корня четвертой степени из единицы удобно взять не первообразный корень $\alpha = \pm i$, а корень $\alpha = -1$. В этом случае резольвента

$$r = x_0 - x_1 + x_2 - x_3$$

принимает при перестановках корней не $4! = 24$ значения, а всего лишь $24/4 = 6$, причем каждое значение принимает ровно четыре раза:

$$r_1 = (x_0 + x_1) - (x_2 + x_3) = -r_4,$$

$$r_2 = (x_0 + x_2) - (x_1 + x_3) = -r_5,$$

$$r_3 = (x_0 + x_3) - (x_1 + x_2) = -r_6.$$

Таким образом, $g(t) = h^4(t)$, где $h(t) = (t^2 - r_1^2)(t^2 - r_2^2)(t^2 - r_3^2)$. Поэтому уравнение $g(t)=0$ эквивалентно уравнению $h(t)=0$, которое является кубическим относительно t^3.

Проводить вычисления резольвент для первообразных корней $\alpha = \pm i$ необходимости нет. В самом деле,

$$(x_0 + x_1 + x_2 + x_3) + r_1 + r_2 + r_3 = 4x_0.$$

Для многочлена пятой степени разрешающее уравнение имеет степень $5! = 120$. Пусть

$$r = x_0 + \alpha x_1 + \alpha^2 x_2 + \alpha^3 x_3 + \alpha^4,$$

где α^5. Если r_1 — одно из значений резольвенты r, то αr_1, $\alpha^2 r_1$, $\alpha^3 r_1$ и $\alpha^4 r_1$ тоже будут значениями резольвенты r. Поэтому разрешающий многочлен $g(t)$ представляет собой многочлен 24-й степени от t^5.

Несмотря на значительные усилия, Лагранжу не удалось решить уравнение $g(t) = 0$. Этот неуспех побудил его считать, что общее уравнение пятой степени нельзя решить в радикалах.

Но для тех уравнений, которые можно решить в радикалах, метод резольвент Лагранжа оказался важным способом их решения. В следующем параграфе мы воспользуемся им для решения в радикалах уравнения $x^n - 1 = 0$.

§ 4. Корни из единицы

Уравнение $x^n = 1$ допускает очевидное решение в радикалах: $x = \sqrt[n]{1}$. Но это решение неудовлетворительно по следующей причине. Поделив многочлен $x^n - 1$ на $x - 1$, получим многочлен

$$f_n(x) = x^{n-1} + x^{n-2} + \ldots + 1.$$

Степень многочлена f_n равна $n - 1$, поэтому формула $x = \sqrt[n]{1}$ дает на один корень больше, чем нужно. Для корней многочлена f_n желательно получить формулу, пользуясь которой, не приходилось бы извлекать корни степени выше $n-1$. В этом параграфе нас будут интересовать лишь такие формулы для корней многочлена f_n.

Разберем сначала несколько примеров. Формулы для корней многочленов $f_3(x) = x^2+x+1$ и $f_4(x) = (x+1)(x^2+1)$ получаются элементарно, поэтому мы сразу перейдем к многочлену $f_5(x) = x^4 + x^3 + x^2 + x + 1$. Положим $y = x + x^{-1}$. Тогда уравнение $f_5(x) = 0$ эквивалентно уравнению

$$y^2 + y - 1 = 0.$$

Корни последнего уравнения вычисляются по формуле

$$y_{1,2} = \frac{-1 \pm \sqrt{5}}{2},$$

а корни уравнений $y_i = x + x^{-1}$ имеют вид

$$\frac{\sqrt{5} - 1 \pm \sqrt{-2\sqrt{5} - 10}}{4}, \qquad \frac{-\sqrt{5} - 1 \pm \sqrt{2\sqrt{5} - 10}}{4}.$$

Число α называют *примитивным корнем n-й степени из единицы*, если $\alpha^n = 1$ и $\alpha^k \neq 1$ при $k = 1, 2, \ldots, n-1$. В этом случае числа $\alpha, \alpha^2, \ldots, \alpha^{n-1}$ попарно различны. Легко проверить, что если α и β — примитивные корни p-й и q-й степени из единицы, причем числа p и q взаимно просты, то $\alpha\beta$ — примитивный корень из единицы степени pq. Поэтому достаточно рассмотреть уравнения $f_m(x) = 0$, где m — степень простого числа.

Для уравнения $f_7(x) = 0$ тоже можно сделать замену $y = x + x^{-1}$. В результате получим уравнение

$$y^3 + y^2 - 2y - 1 = 0.$$

Его корни получим по формуле

$$y_{1,2,3} = \frac{1}{3}\left(-1 + \sqrt[3]{\frac{-7 + 21\sqrt{-3}}{6}} + \sqrt[3]{\frac{-7 - 21\sqrt{-3}}{6}}\right).$$

После этого корни уравнения $f_i(x) = 0$ можно найти, решив квадратные уравнения $x^2 - y_i x + 1 = 0$.

Для многочлена $f_{11} = 0$ замена $y = x + x^{-1}$ не приводит к желаемому результату, поскольку после такой замены получается уравнение 5-й степени, а как его решать, непонятно. Преодолеть возникающие трудности позволяет метод резольвент Лагранжа. Корнями уравнения

$$f_{11}(x) = x^{10} + x^9 + \ldots + x + 1 = 0 \tag{4.1}$$

являются числа $\alpha, \alpha^2, \ldots, \alpha^{10}$ где α — один из примитивных корней 11-й степени из единицы. Степень многочлена f_{11} равна 10, поэтому для построения его резольвенты Лагранжа нужен корень 10-й степени из единицы. Пусть β — примитивный корень

10-й степени из единицы. Упорядочим корни уравнения (4.1) так, что каждый следующий корень является квадратом предыдущего:

$$\alpha, \alpha^2, \alpha^4, \alpha^8, \alpha^5, \alpha^{10}, \alpha^9, \alpha^7, \alpha^3, \alpha^6. \qquad (4.2)$$

Рассмотрим резольвенту Лагранжа, построенную по этой последовательности корней:

$$r_1 = r(\alpha, \beta) = \alpha + \beta\alpha^2 + \beta^2\alpha^4 + \beta^3\alpha^8 + \ldots + \beta^8\alpha^3 + \beta^9\alpha^6.$$

При замене примитивного корня β корнем β^i, $i = 1, \ldots, 10$, получим резольвенты Лагранжа $r_i = r(\alpha, \beta^i)$, $i = 1, \ldots, 10$. Легко проверить, что

$$r_1 + r_2 + \ldots + r_{10} = 10\alpha.$$

Поэтому искомый корень 11-й степени α будет найден, если удастся вычислить величины $r_1, \ldots, r_{10}$.

Покажем теперь, как можно вычислить величины r_i^{10}. Так как $\alpha^{11} = 1$ и $\beta^{10} = 1$, то r_i можно представить в виде

$$\sum_{j,k} a_{ijk}\beta^j\alpha^k, \quad \text{где} \quad 0 \leqslant j \leqslant 9, \quad 0 \leqslant k \leqslant 10,$$

а r_i^{10} можно представить в виде

$$\sum_{j,k} b_{ijk}\beta^j\alpha^k, \quad \text{где} \quad 0 \leqslant j \leqslant 9, \quad 0 \leqslant k \leqslant 10.$$

При этом a_{ijk} и b_{ijk} — целые неотрицательные числа. Запишем r_i^{10} как многочлен от α, причем степени α расположим в порядке (4.2):

$$r_i^{10} = p_{i,0}(\beta) + p_{i,1}(\beta)\,\alpha + p_{i,2}(\beta)\,\alpha^2 + p_{i,3}(\beta)\,\alpha^4 + \ldots + p_{i,10}(\beta)\,\alpha^6. \qquad (4.3)$$

Докажем, что r_i^{10} не зависит от α, т. е. $r_i^{10} = p_i(\beta)$, где p_i — многочлен с целыми коэффициентами. Легко проверить, что $r(\alpha^2, \beta^i) = \beta^{-i} r(\alpha, \beta^i)$, т. е. $r_i(\alpha^2) = \beta^{-i} r_i(\alpha)$. Поэтому $r_i^{10}(\alpha^2) = r_i^{10}(\alpha)$. Таким образом, обозначив для краткости $p_{i,k}(\beta)$ в формуле (4.3) через p_k, получим

$$p_0 + p_1\alpha + p_2\alpha^2 + p_3\alpha^4 + p_4\alpha^8 + \ldots + p_9\alpha^3 + p_{10}\alpha^6 = p_0 + p_1\alpha^2 + p_2\alpha^4 + p_3\alpha^8 + p_4\alpha^5 + \ldots + p_9\alpha^6 + p_{10}\alpha.$$

Следовательно,

$$(p_{10} - p_1)\,\alpha + (p_1 - p_2)\,\alpha^3 + (p_2 - p_3)\,\alpha^4 + \ldots + (p_9 - p_{10})\,\alpha^6 = 0. \qquad (4.4)$$

Можно доказать следующее утверждение.

Теорема 1. *Пусть p — простое число, α и β — примитивные корни степени p и $p-1$ из единицы. Тогда из равенства*

$$q_1(\beta)\,\alpha + q_2(\beta)\,\alpha^2 + \ldots + q_{p-1}(\beta)\,\alpha^{p-1} = 0,$$

где $q_1, q_2, \ldots, q_{p-1}$ — многочлены с целыми коэффициентами, следует, что $q_1(\beta) = q_2(\beta) = \ldots = q_{p-1}(\beta) = 0$.

Такого рода независимость корней α и β кажется почти очевидной. Однако строгое доказательство теоремы 1 не так просто. Оно требует привлечения элементов теории Галуа. Это доказательство мы приводить не будем.

Применив теорему 1 (для $p = 11$) к равенству 4.4, получим $p_{10} = p_1 = p_2 = p_3 = \ldots = p_9$. Следовательно,

$$r_i^{10} = p_0(\beta) + p_1(\beta)\,(\alpha + \alpha^2 + \alpha^3 + \ldots + \alpha^{10}) = p_0(\beta) - p_1(\beta).$$

Подведем итоги. Примитивный корень 10-й степени можно представить в виде произведения примитивных корней пятой и второй степени. Поэтому β можно выразить с помощью квадратных радикалов. Затем можно представить r_i^{10} в виде многочлена от β с целыми коэффициентами. После этого остается воспользоваться формулой

$$\alpha = \frac{1}{10}\left(\sqrt[10]{r_1^{10}} + \ldots + \sqrt[10]{r_{10}^{10}}\right).$$

Эта формула дает, правда, 100 значений, а не 10, как хотелось бы. Нужные 10 значений приходится выбирать среди полученных значений. Поэтому более удобна формула

$$\alpha = \frac{1}{10}\left(r_1 + \frac{r_2 r_1^8}{r_1^8} + \frac{r_3 r_1^7}{r_1^7} + \ldots\right). \qquad (4.5)$$

Дело в том, что величины $r_i r_1^{10-i}$, как и величины r_i^{10}, тоже не зависят от α. Чтобы доказать это, достаточно заметить, что

$$r_i(\alpha^2)\,r_1^{10-i}(\alpha^2) = \beta^{-i} r_i(\alpha)\,(\beta^{-1} r_1(\alpha))^{10-i} = r_i(\alpha)\,r_1^{10-i}(\alpha)$$

Таким образом, формула (4.5) однозначно определяет α после того, как выбрано одно из десяти значений $r_1 = \sqrt[10]{r_1^{10}}$.

Аналогичным образом можно решить и уравнение $f_p(x) = 0$ для любого простого числа $p > 2$. При решении уравнения $f_{11}(x) = 0$ мы воспользовались тем, что примитивные корни 11-й степени из единицы можно упорядочить так, чтобы каждый следующий корень был квадратом предыдущего. Такое расположение корней возможно потому, что числа 2^k, $k = 0, 1, \ldots, 9$, имеют попарно различные остатки по модулю 11. Для остатков по модулю, например, 7, аналогичное утверждение неверно, но для произвольного простого числа p существует такое простое число g, что числа g^k, $k = 0, 1, \ldots, p-2$, имеют попарно различные остатки по модулю p. Иными словами, мультипликативная группа ненулевых вычетов по модулю p является циклической группой, порожденной образующей g. Доказывать это хорошо известное утверждение мы не будем.

Пусть α и β — примитивные корни из единицы степени p и $p-1$ соответственно. Если числа g^k, $k = 0, 1, \ldots, 9$, имеют попарно различные остатки по модулю p, то числа

$$\alpha, \quad \alpha^g, \quad \alpha^{g^2}, \quad \ldots, \quad \alpha^{g^{p-1}}$$

исчерпывают все примитивные корни p-й степени из единицы. Поэтому для многочлена $f_p(x)$ можно рассмотреть резольвенту Лагранжа

$$r_1 = \alpha + \beta\alpha^g + \beta^2\alpha^{g^2} + \ldots + \beta^{p-1}\alpha^{g^{p-1}}.$$

Положим $r_i = r_1(\alpha, \beta^i)$. Легко проверить, что $r_1(\alpha^g, \beta^i) = \beta^{-1}c_1(\alpha, \beta^i)$. Следовательно, величины r_i^{p-1} и $r_i r^{p-1-i}$ не изменяются при замене α на α^g. Как и в случае $p = 11$, с помощью этого свойства и теоремы 1, можно доказать, что величины r_i^{p-1} и $r_i r_1^{p-1-i}$ представляют собой многочлены от β с целыми коэффициентами. Для β выражение в радикалах можно получить по индукции, так как β — корень из единицы степени $p-1 < p$. Формула для вычисления α имеет вид

$$\alpha = \frac{1}{p-1}\left(r_1 + \frac{r_2 r_1^{p-3}}{r_1^{p-3}} + \frac{r_3 r_1^{p-4}}{r_1^{p-4}} + \ldots\right).$$

Эта формула дает однозначное выражение для α после того, как выбрано одно из значений $r_1 = \sqrt[p-1]{r_1^{p-1}}$.

§ 5. Теорема Абеля о неразрешимости в радикалах общего уравнения пятой степени

Работы Лагранжа побудили многих геометров, а в то время так называли всех математиков, заняться поисками доказательства неразрешимости в радикалах общего уравнения пятой степени и уравнений более высоких степеней. В 1798–1813 гг. появилось несколько работ итальянского математика Паоло Руффини (1765–1822). Вслед за Лагранжем он рассматривал подстановки корней уравнений и ввел при этом термин «группа перестановок». Эта серия работ [А9] увенчалась доказательством теоремы о неразрешимости в радикалах уравнений пятой и более высоких степеней. К сожалению, в этом доказательстве был существенный пробел. Руффини без основания предполагал, что радикалы рационально выражаются через корни исходного уравнения (см. теорему 4).

Первое полное доказательство теоремы о неразрешимости общего уравнения пятой степени получил гениальный норвежский математик Нильс Хенрик Абель (1802–1829). Это доказательство он изложил в мемуаре «Доказательство невозможности алгебраического решения общих уравнений пятой степени», опубликованном в первом номере журнала Крелля за 1825 г.

Будем говорить, что уравнение

$$F(x) = x^n + c_1 x^{n-1} + \ldots + c_n = 0$$

является *общим уравнением* n-й степени, если его коэффициенты $c_1, \ldots, c_n$ рассматриваются как независимые переменные над некоторым полем L. В дальнейшем будем считать, что $L = \mathbb{Q}$.

Присоединив коэффициенты $c_1, \ldots, c_n$ к полю $\mathbb{Q}$, получим поле $\Delta = \mathbb{Q}(c_1, \ldots, c_n)$. Его называют *полем рациональности* данного уравнения.

Присоединив к полю Δ корни $\alpha_1, \ldots, \alpha_n$ уравнения (5.1), получим поле $\Delta(F)$. Его называют *нормальным полем* уравнения (5.1) или *полем Галуа* этого уравнения.

Будем говорить, что уравнение (5.1) *разрешимо в радикалах*, если поле $\Delta(F)$ содержится в расширении R поля Δ, полученном путем присоединения к Δ некоторых радикалов $\rho_1 = \sqrt[s_1]{a_1}$, $\rho_2 = \sqrt[s_2]{a_2}, \ldots, \rho_m = \sqrt[s_1]{a_m}$, где $a_1 \in \Delta$, $a_2 \in \Delta(\rho_1)$, $a_3 \in \Delta(\rho_1, \rho_2), \ldots, a_m \in \Delta(\rho_1, \ldots, \rho_{m-1})$.

Пример. Пусть $F(x) = x^2 + c_1 x + c_2$. Тогда $\Delta = \mathbb{Q}(c_1, c_2)$ и $\Delta(F) = \Delta\left(\sqrt{a_1}\right)$, где $a_1 = c_1^2 - 4c_2 \in \Delta$.

Отметим, что показатели $s_1, \ldots, s_v$ радикалов $\rho_1, \rho_2, \ldots, \rho_m$ можно считать простыми числами. В самом деле, если $s_k = uv$, то присоединение радикала $\rho_k = \sqrt[s_k]{a_k}$ можно заменить последовательным присоединением радикалов $\rho = \sqrt[u]{a_k}$ и $\rho_1 = \sqrt[v]{\rho}$. Поэтому в дальнейшем будем рассматривать лишь присоединение радикалов с простыми показателями.

Предположим, что уравнение (5.1) разрешимо в радикалах. Присоединим к полю Δ первообразные корни из единицы $\varepsilon_1, \ldots, \varepsilon_m$, степени которых равны степеням радикалов $\rho_1, \ldots, \rho_m$ соответственно. Полученное поле обозначим K. Так как $\Delta \subset K$, то $\Delta(F) \subset \Delta(\rho_1, \ldots, \rho_m) \subset K(\rho_1, \ldots, \rho_m)$

Для доказательства теоремы Абеля нам понадобятся некоторые вспомогательные утверждения.

Теорема 1. *Пусть p — простое число, k — поле нулевой характеристики. Многочлен $x^p - a$ приводим над полем k тогда и только тогда, когда $a = b^p$ для некоторого $b \in k$.*

Доказательство. Предположим, что $x^p - a = f(x)\,g(x)$, где $f(x)$ и $g(x)$ — многочлены над полем k. Пусть ε — примитивный корень p-ой степени из единицы и $\beta = \sqrt[p]{a}$. Тогда

$$f(x) = x^r + c_1 x^{r-1} + \ldots + c_r = (x - \varepsilon^{n_1}\beta)\ldots(x - \varepsilon^{n_r}\beta).$$

Поэтому $\pm\varepsilon^l\beta^r = c^r \in k$, где $l = n_1 + \ldots + n_r$. А так как $(\varepsilon^l)^p = 1$, то $(\pm\beta^r)^p = (c_r)^p$, т. е. $a^r = (\pm c_r)^p$. Число p простое и $1 \leqslant r = \deg f < p$, поэтому $rs + pt = 1$ для некоторых чисел s и t. Следовательно, $a = a^{rs}a^{pt} = (\pm c_r^s a^t)^p = b^p$, где $b = \pm c_r^s a^t \in k$.

Ясно так же, что если $a = b^p$, то многочлен $x^p - a$ приводим, так как он делится на $x - b$. □

Теорема 2. *Пусть s — простое число и $a_i \in k = K(\rho_1, \ldots, \rho_{i-1})$. Если $\rho_i = \sqrt[s]{a_i} \notin k$, то $\rho_i^l \in k$ тогда и только тогда, когда l делится на s.*

Доказательство. Если $l = ns$, то $\rho_i^l = a_i^n \in k$, так как $a_i \in k$.

Предположим теперь, что $\rho_i^l = a \in k$, причем $l = sq + r$, где $0 < r < s$. Тогда $a = \rho_i^l = (a_i)^q \rho_i^r$, а значит, $\rho_i^r = b$, где $b = a(a_i)^{-q}$. Над полем k многочлены $x^s - a_i$ и $x^r - b$ имеют общий корень ρ_i, поэтому они имеют общий делитель, степень которого не превосходит $r < s$. В частности, многочлен $x^s - a_i$ приводим над полем k. Из теоремы 1 следует, что $a_i = b^s$, где $b \in k$. Ясно, что $b = \varepsilon\rho_i$, где ε — корень степени s из единицы. А так как $\varepsilon \in K \subset k$, то $\rho_i \in k$. Получено противоречие. □

Можно считать, что $\rho_1, \ldots, \rho_m$ — минимальная последовательность радикалов с простыми степенями, требуемая для вычисления корня α уравнения (5.1), т. е. любая такая последовательность содержит по крайней мере m таких радикалов. При этом условии справедливо следующее утверждение.

Теорема 3. *Корень α уравнения* (5.1) *можно представить в виде*

$$\alpha = u_0 + \rho + u_2\rho^2 + \ldots + u_{s-1}\rho^{s-1}$$

где s — степень радикала ρ_m, $\rho = \sqrt[s]{a}$, где a и все u_l — какие-то элементы поля $k = K(\rho_1, \ldots, \rho_{m-1})$.

Доказательство. Так как $\alpha \in K(\rho_1, \ldots, \rho_m) = k(\rho_m)$ и $\rho_m^s \in k$, то

$$\alpha = b_0 + b_1\rho_m + b_2\rho_m^2 + \ldots + b_{s-1}\rho_m^{s-1}, \tag{5.2}$$

где $b_i \in k$. Трудность заключается лишь в том, чтобы добиться равенства $b_1 = 1$. По условию $\alpha \notin k$, поэтому хотя бы одно из чисел $b_1, \ldots, b_{s-1}$ отлично от нуля. Пусть $b_l \neq 0$, где $1 \leqslant l \leqslant s$. Положим $\rho = b_l\rho_m^l$. Так как число s простое, то $ul + vs = 1$ для некоторых целых чисел u и v. При этом $\rho^u = b_l^u\rho_m^{ul} = b_l^u\rho_m^{1-vs} = b_l^u a^{-v}\rho_m$, т. е. $\rho_m = c\rho^u$, где $c = b_l^{-u}a^v \in k$. Так как $\rho_m \notin k$, то $\rho \in k$. Ясно так же, что $\rho^s b_l^s \rho_m^{ls} = b_l^s a^l \in k$.

Заменим в выражении (5.2) ρ_m на $c\rho^u$, вспомнив при этом, что $b_l\rho_m^l = \rho$. В результате получим

$$\alpha = b_0 + b_1 c\rho^u + b_2 c^2\rho^{2u} + \ldots + \rho + \ldots + b_{s-1}\rho^{(s-1)u} \tag{5.3}$$

Из теоремы следует, что $\rho^t \in k$ тогда и только тогда, когда t делится на s. А так как числа u и s взаимно простые, то элементы $1, \rho^u, \rho^{2u}, \ldots, \rho^{(s-1)u}$ линейно независимы над полем k, причем набор этих элементов совпадает с набором $1, \rho, \rho^2, \ldots, \rho^{s-1}$ (в другом порядке). Таким образом, (5.3) дает требуемое выражение для α:

$$\alpha = b_0 + \rho + b_2'\rho^2 + \ldots + b_{s-1}'\rho^{s-1}. \qquad \square$$

Теорема 4. *Минимальную последовательность радикалов $\rho_1, \ldots, \rho_m$ для вычисления корня α многочлена* (5.1) *можно выбрать так, что $\rho_1, \ldots, \rho_m$ представляют собой полиномы над K от корней $\alpha_1, \ldots, \alpha_m$ многочлена* (5.1).

Доказательство. Будем исходить из произвольной минимальной последовательности $\rho_1, \ldots, \rho_m$. Напомним, что через K мы обозначили поле, полученное в результате присоединения к полю Δ первообразных корней $\varepsilon_1, \ldots, \varepsilon_m$, степени которых равны степеням радикалов $\rho_1, \ldots, \rho_m$. Согласно теореме 3 радикал ρ_m можно заменить радикалом ρ той же самой степени s так, что

$$\alpha = u_0 + \rho + u_2\rho^2 + \ldots + u_{s-1}\rho^{s-1},$$

где $u_l \in k = K(\rho_1, \ldots, \rho_{m-1})$ и $\rho^s = a \in k$. Покажем, что для любого корня ξ многочлена $x^s - a$ величина

$$\alpha(\xi) = u_0 + \xi + u_2\xi^2 + \ldots + u_{s-1}\xi^{s-1}$$

является корнем многочлена (5.1). Подставим в многочлен $F(x) = x^n + c_1x^{n-1} + \ldots + c_n$ вместо x величину $\alpha(\xi)$. Учитывая, что $\xi^s = a \in k$, в результате получим выражение вида

$$b_0 + b_1\xi + \ldots + b_{s-1}\xi^{s-1},$$

где $b_l \in k$. Многочлены $x^s - a$ и $b_0 + b_1x + \ldots + b_{s-1}x^{s-1}$ имеют общий корень ρ, поэтому они имеют общий делитель над k. Но согласно теореме 1 многочлен $x^s - a$ неприводим над k, поэтому $b_0 = b_1 = \ldots = b_{s-1} = 0$. Это означает, что если ξ — корень многочлена $x^s - a$, то $\alpha(\xi)$ — корень многочлена (5.1). Пусть ε — примитивный корень степени s из единицы. Тогда $\xi = \varepsilon^r\rho$, поэтому величины

$$\alpha_{r+1} = u_0 + \varepsilon^r\rho + u_2\varepsilon^{2r}\rho^2 + \ldots + u_{s-1}\varepsilon^{(s-1)r}\rho^{s-1}$$

при $r = 0, 1, \ldots, s-1$ будут корнями многочлена (5.1). Например, для $s = 3$ получим

$$\alpha_1 = u_0 + \rho + u_2\rho^2, \quad \alpha_2 = u_0 + s\rho + u_2\varepsilon^2\rho^2, \quad \alpha_3 = u_0 + \varepsilon^2\rho + u_2\varepsilon\rho^2.$$

А так как $1 + \varepsilon + \varepsilon^2 = 0$, то

$$\alpha_1+\alpha_2+\alpha_3=3u_0, \quad \alpha_1+\varepsilon^{-1}\alpha_2+\varepsilon^{-2}\alpha_3=3\rho, \quad \alpha_1+\varepsilon^{-2}\alpha_2+\varepsilon^{-1}\alpha_3=3u_2\rho^2.$$

Таким образом, $\rho = \frac{1}{3}(\alpha_1 + \varepsilon^2\alpha_2 + \varepsilon\alpha_3)$. Для $s > 3$ формулы получаются более громоздкие, но рассуждения те же самые. Доказательство теоремы для последнего радикала ρ_m завершено. Перейдем к радикалу ρ_{m-1}.

Выше было показано (для $s = 3$), что величины $u_0, \rho, u_2\rho^2, \ldots, u_{s-1}\rho^{s-1}$ полиномиально выражаются через корни $\alpha_1, \ldots, \alpha_m$ многочлена (5.1). Кроме того, они лежат в поле $K(\rho_1, \ldots, \rho_{m-1})$, поэтому каждую из указанных величин можно представить в виде

$$\nu_0 + \nu_1\rho_{m-1} + \nu_2\rho_{m-1}^2 + \ldots + \nu_{l-1}\rho_{m-1}^{l-1},$$

где $\nu_l \in K(\rho_1, \ldots, \rho_{m-2})$. Последовательность радикалов $\rho_1, \ldots, \rho_m$ минимальна, поэтому равенства $\nu_1 = \nu_2 = \ldots = \nu_{l-1} = 0$ не могут выполняться сразу для всех величин, так как иначе радикал ρ_{m-1} можно было бы исключить. Поэтому существует соотношение вида

$$\nu_0 + \nu_1\rho_{m-1} + \nu_2\rho_{m-1}^2 + \ldots + \nu_{l-1}\rho_{m-1}^{l-1} = r(\alpha_1, \ldots, \alpha_n)$$

где $\nu_l \in K(\rho_1, \ldots, \rho_{m-2})$, причем не все элементы $\nu_1, \ldots, \nu_{l-1}$ равны нулю, а $r(\alpha_1, \ldots, \alpha_n)$ — некоторый многочлен над полем K. Рассмотрим многочлен

$$G(x) = \prod (x - r(\alpha_{\sigma(1)}, \ldots, \alpha_{\sigma(n)})),$$

где произведение берется по всем подстановкам $\sigma \in S_n$. Коэффициенты многочлена G являются симметрическими многочленами от корней многочлена (5.1), поэтому они полиномиально выражаются через коэффициенты многочлена (5.1). Таким образом, G — многочлен над K, причем он имеет корень

$$\beta = \nu_0 + \nu_1\rho_{m-1} + \nu_2\rho_{m-1}^2 + \ldots + \nu_{l-1}\rho_{m-1}^{l-1}$$

Ясно так же, что корень β выражается в радикалах (с помощью последовательности радикалов $\rho_1, \ldots, \rho_{m-1}$). Согласно теореме 3, заменив радикал ρ_{m-1} радикалом ρ' той же самой степени, можно добиться равенства $\nu_1 = 1$. К радикалу ρ' теперь можно применить те же самые рассуждения, которые мы применили к радикалу ρ. Затем переходим к радикалу ρ_{m-2} и т. д. до радикала ρ_1. □

Теперь можно приступить непосредственно к доказательству теоремы Абеля.

Теорема 5 (Абель). *При $n \geqslant 5$ корень общего многочлена степени n нельзя выразить в радикалах.*

Доказательство. Предположим, что некоторый корень α_1 общего многочлена степени n

$$x^n + c_1 x^{n-1} + c_2 x^{n-2} + \ldots + c_n$$

можно выразить в радикалах. Тогда согласно теоремам 1–4 для α_1 существует выражение в радикалах следующего специального вида. Корень α_1 получается последовательным присоединением радикалов $\rho_1, \ldots, \rho_m$ простых степеней, а эти радикалы, в свою очередь, являются полиномами от корней $\alpha_1, \ldots, \alpha_n$ исходного многочлена. Точнее говоря, пусть $\varepsilon_1, \ldots, \varepsilon_m$ — примитивные корни из единицы, степени которых равны степеням радикалов $\rho_1, \ldots, \rho_{m-2}$ соответственно, $\Delta = \mathbb{Q}(c_1, \ldots, c_n)$ и

$$K = \Delta(\varepsilon_1, \ldots, \varepsilon_m) = \mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m, c_1, \ldots, c_n).$$

Тогда α_1 полиномиально выражается через $\rho_1, \ldots, \rho_m$ над полем K, т. е.

$$\alpha_1 = r(\rho_1, \ldots, \rho_m, c_1, \ldots, c_n),$$

где r — полином над $\mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m)$. В свою очередь, радикалы $\rho_1, \ldots, \rho_m$ полиномиально выражаются над полем K через $\alpha_1, \ldots, \alpha_n$, т. е.

$$\rho_1 = r_1(\alpha_1, \ldots, \alpha_n, c_1, \ldots, c_n),$$

где r_1 — полином над $\mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m)$. Так как мы имеем дело с общим многочленом степени n, то можно считать, что $\alpha_1, \ldots, \alpha_n$ — независимые переменные, а $c_1, \ldots, c_n$ представляют собой (с точностью до знака) элементарные симметрические многочлены от $\alpha_1, \ldots, \alpha_n$.

Покажем, что при $n \geqslant 5$ предположение о разрешимости в радикалах общего алгебраического уравнения степени n приводит к противоречию. Рассмотрим для этого перестановку

$$T = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 2 & 3 & 4 & 5 & 1 & 6 & \ldots & n \end{pmatrix},$$

которая циклически переставляет первые 5 элементов, а остальные элементы оставляет неподвижными. Докажем, что при действии T на корни $\alpha_1, \ldots, \alpha_n$ первый радикал ρ_1 не изменяется. Так как

$$\rho_1 = r_1(\alpha_1, \ldots, \alpha_n, c_1, \ldots, c_n) = \sqrt[p]{a_1},$$

где a_1 — многочлен от $c_1, \ldots, c_n$ над полем $\mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m)$, то равенство $\rho_1^p = a_1$ можно рассматривать как соотношение вида

$$\varphi(\alpha_1, \ldots, \alpha_n, c_1, \ldots, c_n) = 0,$$

где φ — многочлен над полем $\mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m)$. Покажем, что соотношение такого вида сохраняется при любой перестановке корней $\alpha_1, \ldots, \alpha_n$. Пусть $\beta_1 = \alpha_{i_1}, \ldots, \beta_n = \alpha_{i_n}$, где $i_1, \ldots, i_n$ — некоторая перестановка чисел $1, 2, \ldots, n$. Тогда

$$\varphi(\beta_1, \ldots, \beta_n, d_1, \ldots, d_n) = 0,$$

где $d_i = c_i(\beta_1, \ldots, \beta_n)$. А так как функции c_i симметрические, то $d_i = c_i(\alpha_1, \ldots, \alpha_n) = c_i$. Поэтому

$$\varphi(\alpha_{i_1}, \ldots, \alpha_{i_n}, c_1, \ldots, c_n) = 0.$$

Итак, соотношение $\rho_1^p = a_1$ сохраняется при действии перестановки T на корни $\alpha_1, \ldots, \alpha_n$, т. е. $T(\rho_1^p) = T(a_1)$. Ясно, что $T(\rho_1^p) = T(\rho_1)^p$. А так как a_1 зависит лишь от симметрических функций корней, то $T(a_1) = a_1$. Следовательно, $T(\rho_1) = \varepsilon_1^\lambda \rho_1$ и $T^m(\rho_1) = \varepsilon_1^{m\lambda}\rho_1$. Но $T^5 - I$ — тождественная перестановка, поэтому $\varepsilon_1^{5\lambda}\rho_1 = T^5(\rho_1) = \rho_1$, т. е. $\varepsilon_1^{5\lambda} = 1$.

Обратимся теперь к перестановкам

$$U = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 1 & 2 & 4 & 5 & 3 & 6 & \ldots & n \end{pmatrix}$$

и

$$V = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & \ldots & n \\ 2 & 3 & 1 & 4 & 5 & 6 & \ldots & n \end{pmatrix}.$$

Легко проверить, что $U^3 = V^3 = I$, поэтому $U(\rho_1) = \varepsilon_1^\mu \rho_1$ и $V(\rho_1) = \varepsilon_1^\nu \rho_1$, причем $\varepsilon_1^{3\mu} = \varepsilon_1^{3\nu} = 1$. Кроме того, $VU = T$, поэтому

$$T(\rho_1) = VU(\rho_1) = \varepsilon_1^{\mu+\nu}\rho_1$$

Следовательно, $\varepsilon_1^\lambda = \varepsilon_1^{\mu+\nu}$, а значит, $\varepsilon_1^\lambda = \varepsilon_1^{6\lambda}\varepsilon_1^{-5\lambda} = \varepsilon_1^{6(\mu+\nu)} = 1$, так как $\varepsilon_1^{5\lambda} = \varepsilon_1^{6\mu} = \varepsilon_1^{6\nu} = 1$. В итоге получаем $T(\rho_1) = \rho_1$.

Переходя последовательно к радикалам $\rho_2, \ldots, \rho_m$, аналогично получим $T(\rho_i) = \rho_i$ при $i = 2, \ldots, m$.

Так как $\rho_i = r_i(\alpha_1, \ldots, \alpha_n, c_1, \ldots, c_n)$, то равенство

$$\alpha_1 = r(\rho_1, \ldots, \rho_m, c_1, \ldots, c_n)$$

можно рассматривать как соотношение между $\alpha_1, \ldots, \alpha_n, c_1, \ldots, c_n$ над полем $\mathbb{Q}(\varepsilon_1, \ldots, \varepsilon_m)$. Это соотношение сохраняется под действием подстановки T, т. е.

$$T(\alpha_1) = r(T(\rho_1), \ldots, T(c_n)) = r(\rho_1, \ldots, c_n),$$

так как $T(c_i) = c_i$ и $T(\rho_i) = \rho_i$. Таким образом, $T(\alpha_1) = \alpha_1$. С другой стороны, согласно определению T получаем $T(\alpha_1) = \alpha_2$, а значит, $\alpha_1 = \alpha_2$. Равенство $\alpha_1 = \alpha_2$ противоречит независимости корней общего уравнения. □

§ 6. Преобразование Чирнгауза. Уравнение пятой степени в форме Бринга

В 1683 г. в лейпцигском журнале «Acta eruditorum» Э. В. фон Чирнгауз (1651–1708) опубликовал способ преобразования алгебраических уравнений, который, как ему казалось, позволял решить в радикалах уравнение любой степени [A12]. Лейбниц сразу же опроверг заявление Чирнгауза о всемогуществе его преобразования. Дело в том, что при решении уравнений 5-й степени с помощью преобразования Чирнгауза приходится решать уравнение 24-й степени. Несмотря на это, преобразование Чирнгауза имеет важные приложения. Например, с его помощью любое уравнение пятой степени без кратных корней можно привести к виду $y^5 + 5y = a$, решая при этом лишь уравнения второй и третьей степени. А в главе 2 мы покажем, что уравнения такого вида можно решать с помощью тэта-функций.

Преобразование Чирнгауза уравнения $x^n + c_1 x^{n-1} + \ldots + c_n$ заключается в следующем. Пусть $x_1, \ldots, x_n$ — корни этого уравнения. Рассмотрим рациональную функцию φ, которая не обращается в бесконечность в точках $x_1, \ldots, x_n$. Положим $y_i = \varphi(x_i)$ и найдем уравнение

$$y^n + q_1 y^{n-1} + \ldots + q_n = 0$$

корнями которого являются $y_1, \ldots, y_n$. Далее мы покажем, что если это уравнение не имеет кратных корней, то x_i можно выразить через y_i. Выбирая подходящим образом функцию φ, можно добиться того, чтобы коэффициенты $q_1, \ldots, q_{n-1}$ стали равны нулю. Но для этого потребуется решить уравнение степени $(n-1)!$; именно на это обстоятельство указал Лейбниц.

Не теряя общности, в качестве рациональной функции φ можно взять многочлен степени не более $n-1$. Дело в том, что справедливо следующее утверждение.

Теорема 1. *Пусть $x_1, \ldots, x_n$ — корни многочлена f степени n и $\varphi = P/Q$, где P и Q — многочлены, причем $Q(x_i) \neq 0$, $i = 1, \ldots, n$. Тогда существует многочлен g степени не более $n-1$, значения которого в точках $x_1, \ldots, x_n$ совпадают со значениями функции φ в этих точках.*

Доказательство. По условию многочлены f и Q не имеют общих корней, поэтому они взаимно просты. Следовательно, существуют многочлены a и b, для которых $af + bQ = 1$. Так как $f(x_i) = 0$, то $b(x_i) = 1/Q(x_i)$. Поэтому

$$\varphi(x_i) = P(x_i)/Q(x_i) = P(x_i)\, b(x_i)$$

Таким образом, в качестве требуемого многочлена g можно взять остаток от деления многочлена Pb на многочлен f. □

В дальнейшем будем считать, что к уравнению

$$f(x) = x^n + c_1 x^{n-1} + \ldots + c_n$$

применяется преобразование

$$y = g(x) = p_0 + p_1(x) + \ldots + p_{n-1} x^{n-1}.$$

Покажем, как в этом случае можно вычислить коэффициенты многочлена $y^n + q_1 y^{n-1} + \ldots + q_n$ с корнями $y_i = g(x_i)$, $i = 1, \ldots, n$.

Чтобы избежать громоздких обозначений, ограничимся случаем $n = 3$. Если $x^3 = -c_1 x^2 - c_2 x - c_3$, то

$$yx = p_0 x + p_1 x^2 + p_2(-c_1 x^2 - c_2 x - c_3) = p_0' + p_1' x + p_2' x^2.$$

Аналогично $yx^2 = p_0'' + p_1'' x + p_2'' x^2$, причем p_i'' — линейные функции параметров p_i. Таким образом, если x_i — корень многочлена f и $y_i = g(x_i)$, то система уравнений

$$\begin{cases} (p_0 - y)\, z_0 + p_1 z_1 + p_2 z = 0, \\ p_0' z_0 + (p_1' - y)\, z_1 + p_2' z = 0, \\ p_0'' z_0 + p_1'' z_1 + (p_2'' - y)\, z = 0. \end{cases} \tag{6.1}$$

имеет ненулевое решение $(z_0, z_1, z_2) = (1, x_i, x_i^2)$. Положим

$$A = \begin{pmatrix} p_0 & p_1 & p_2 \\ p_0' & p_1' & p_2' \\ p_0'' & p_1'' & p_2'' \end{pmatrix}.$$

Тогда $\det(A - yE) = 0$ для $y_i = g(x_i)$. В том случае, когда многочлен $\det(A - yE)$ не имеет кратных корней, он совпадает с искомым многочленом $y^n + q_1 y^{n-1} + \ldots + q_n$. Так как элементы матрицы A линейно зависят от параметров p_i, то коэффициент q_k является многочленом степени k от параметров p_i.

В том случае, когда многочлен $y^n + q_1 y^{n-1} + \ldots + q_n$ не имеет кратных корней, матрица A не имеет кратных собственных значений. Поэтому каждому собственному значению матрицы A соответствует единственное с точностью до пропорциональности решение системы (6.1). Это означает, что по корню y_i преобразованного многочлена однозначно восстанавливается корень x_i исходного многочлена, причем x_i рационально выражается через y_i.

Преобразование Чирнгауза позволяет решить в радикалах уравнения третьей и четвертой степени. Кубическое уравнение можно привести к виду

$$y^3 + q_3 = 0,$$

решив систему из линейного уравнения $q_1 = 0$ и уравнения второй степени $q_2 = 0$, зависящих от параметров p_0, p_1, p_2. Для этого нужно решить квадратное уравнение.

Уравнение четвертой степени можно привести к виду

$$y^4 + q_2 y^2 + q_4 = 0.$$

Для этого нужно решить систему из линейного уравнения $q_1 = 0$ и уравнения третьей степени $q_3 = 0$, что сводится к решению кубического уравнения.

Уравнение пятой степени можно привести к виду

$$y^5 + q_4 y + q_5 = 0,$$

решив систему уравнений $q_1 = q_2 = q_3 = 0$. Для этого нужно решить уравнение шестой степени. Более тонкий анализ, проведенный в 1789 г. шведским математиком Брингом, показывает, что в

данном случае вместо уравнения шестой степени достаточно решить уравнения второй и третьей степени, поступив следующим образом. Чтобы удовлетворить условию $q_1 = 0$, выразим один из параметров $p_0, \ldots, p_4$ как линейную функцию остальных параметров. Тогда коэффициент q_2 будет представлять собой квадратичную форму относительно четырех из параметров p_i. Эту квадратичную форму можно привести к виду

$$u_1^2 + u_2^2 - v_1^2 - v_2^2,$$

где u_j и v_j — линейные функции от p_i (для этого понадобится извлекать квадратные корни). Чтобы удовлетворить равенству $q_2 = 0$, достаточно решить систему линейных уравнений $u_1 = v_1$, $u_2 = v_2$. После этого остается два параметра, причем относительно них уравнение q_3 представляет уравнение третьей степени. В итоге получаем уравнение вида $y^5 + q_4 y + q_5 = 0$. В том случае, когда $q_4 \neq 0$, с помощью линейной замены это уравнение можно привести к виду $y^5 + 5y = a$.

§ 7. Уравнения пятой степени, разрешимые в радикалах

Согласно теореме Абеля корни уравнения пятой степени в общем случае не могут быть выражены в виде алгебраических функций от своих коэффициентов. Поэтому весьма интересно найти условия на коэффициенты, при выполнении которых уравнение разрешимо в радикалах, а также получить в явном виде корни этих разрешимых уравнений. Для уравнения общего вида такие условия были найдены английским математиком А. Кэли [В10]. Ввиду сложности результата мы не станем его приводить, отсылая читателя или к мемуару самого Кэли или к учебнику Д. А. Граве [Б13], а сами ограничимся случаем, когда уравнение задано в упрощенной форме

$$x^5 + ax + b = 0, \tag{7.1}$$

где a и b — рациональные числа и многочлен в левой части неприводим. (Для приводимого многочлена уравнение пятой степени очевидно разрешимо в радикалах, поскольку оно распадается в два уравнения степеней, не превосходящих четырех.)

Чтобы описать разрешимые уравнения пятой степени (7.1), мы обобщим метод Кардано решения кубических уравнений вида

$x^3+ax+b=0$. Прежде всего, переформулируем этот метод так, чтобы его можно было применить и к уравнениям (7.1).

Пусть u_1, u_2 — комплексные числа и w — примитивный кубический корень из единицы. Раскрывая скобки в произведении

$$\left(x-(u_1+u_2)\right)\left(x-(wu_1+w^2u_2)\right)\left(x-(w^2u_1+wu_2)\right), \qquad (7.2)$$

мы получим многочлен

$$x^3-3u_1u_2x-(u_1^3+u_2^3). \qquad (7.3)$$

Так как

$$x_j=w^ju_1+w^{2j}u_2, \qquad j=0,1,2,$$

являются корнями кубического многочлена (7.2), то подставляя их в (7.3), мы получим три тождества:

$$(w^ju_1+w^{2j}u_2)^3-3u_1u_2(w^ju_1+w^{2j}u_2)-(u_1^3+u_2^3)=0.$$

Таким образом, кубическое уравнение $x^3+ax+b=0$ имеет три решения

$$x_j=w^ju_1+w^{2j}u_2, \qquad j=0,1,2,$$

где u_1^3 и u_2^3 находятся из системы уравнений

$$\begin{cases} u_1^3+u_2^3=-b, \\ u_1^3u_2^3=-\left(\dfrac{a}{3}\right)^3. \end{cases}$$

Очевидным обобщением многочлена (7.2) является многочлен пятой степени

$$\prod_{j=0}^{4}\left(x-(w^ju_1+w^{4j}u_2)\right), \qquad (7.4)$$

где теперь w — примитивный корень пятой степени из единицы. Раскрывая это произведение, мы получим, что многочлен пятой степени

$$x^5+ax^3+\frac{a^2}{5}x+b$$

(называемый иногда многочленом Муавра) имеет корни

$$x_j=w^ju_1+w^{4j}u_2, \qquad j=0,1,2,3,4,$$

где u_1^5 и u_2^5 определяются системой уравнений

$$\begin{cases} u_1^5+u_2^5=-b, \\ u_1^5u_2^5=-\left(\dfrac{a}{5}\right)^5. \end{cases}$$

Поскольку нас интересует не многочлен Муавра, а многочлен

$$x^5+ax+b,$$

рассмотрим вместо (7.4) многочлен пятой степени

$$\prod_{j=0}^{4}\left(x-(w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4)\right), \qquad (7.5)$$

где u_1, u_2, u_3, u_4 — ненулевые вещественные числа и w — примитивный корень пятой степени из единицы. Записав (7.5) в виде многочлена по x и подставив в этот многочлен величины

$$x_j=w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4,$$

мы получим для $j=0,1,2,3,4$ следующие тождества:

$$\begin{aligned}(w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4)^5-&\\ -5U(w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4)^3-&\\ -5V(w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4)^2-&\\ -5W(w^ju_1+w^{2j}u_2+w^{3j}u_3+w^{4j}u_4)+&\\ +5(X-Y)-Z=0,& \qquad (7.6)\end{aligned}$$

где

$$\begin{aligned} U&=u_1u_4+u_2u_3,\\ V&=u_1u_2^2+u_2u_4^2+u_3u_1^2+u_4u_3^2,\\ W&=u_1^2u_4^2+u_2^2u_3^2-u_1^3u_2-u_2^3u_4-u_3^3u_1-u_4^3u_3-u_1u_2u_3u_4,\\ X&=u_1^3u_3u_4+u_2^3u_1u_3+u_3^3u_2u_4+u_4^3u_1u_2,\\ Y&=u_1u_3^2u_4^2+u_2u_1^2u_3^2+u_3u_2^2u_4^2+u_4u_1^2u_2^2,\\ Z&=u_1^5+u_2^5+u_3^5+u_4^5. \end{aligned}$$

Наша цель — найти вещественные числа u_1, u_2, u_3, u_4, удовлетворяющие уравнениям

$$u_1u_4 + u_2u_3 = 0, \tag{7.7}$$

$$u_1u_2^2 + u_2u_4^2 + u_3u_1^2 + u_4u_3^2 = 0, \tag{7.8}$$

$$5\,(u_1^2u_4^2 + u_2^2u_3^2 - u_1^3u_2 - u_2^3u_4 - u_3^3u_1 - u_4^3u_3 - u_1u_2u_3u_4) = a, \tag{7.9}$$

$$5[(u_1^3u_3u_4 + u_2^3u_1u_3 + u_3^3u_2u_4 + u_4^3u_1u_2) - (u_1u_3^2u_4^2 + u_2u_1^2u_3^2 + u_3u_2^2u_4^2 + u_4u_1^2u_2^2)] - (u_1^5 + u_2^5 + u_3^5 + u_4^5) = b. \tag{7.10}$$

Для таких u_1, u_2, u_3, u_4 многочлен (7.5) принимает вид x^5+ax+b и имеет корни

$$x_j = w^j u_1 + w^{2j} u_2 + w^{3j} u_3 + w^{4j} u_4, \quad j = 0,1,2,3,4. \tag{7.11}$$

Теорема 1. *Пусть a и b — такие рациональные числа, что многочлен пятой степени $x^5 + ax + b$ неприводим. Тогда уравнение*

$$x^5 + ax + b = 0$$

разрешимо в радикалах в том и только том случае, когда существует такие рациональные числа числа $\varepsilon = \pm 1$, $c \geqslant 0$ и $e \neq 0$, что

$$a = \frac{5e^4(3 - 4\varepsilon c)}{c^2 + 1}, \quad b = \frac{-4e^5(11\varepsilon + 2c)}{c^2 + 1}. \tag{7.12}$$

В этом случае корни уравнения имеют вид

$$x_j = e(w^j u_1 + w^{2j} u_2 + w^{3j} u_3 + w^{4j} u_4), \quad j = 0,1,2,3,4. \tag{7.13}$$

где $w = e^{2\pi i/5}$ и

$$u_1 = \left(\frac{v_1^2 v_3}{D^2}\right)^{1/5}, \quad u_2 = \left(\frac{v_3^2 v_4}{D^2}\right)^{1/5}, \quad u_3 = \left(\frac{v_2^2 v_1}{D^2}\right)^{1/5}, \quad u_4 = \left(\frac{v_4^2 v_2}{D^2}\right)^{1/5}, \tag{7.14}$$

$$\begin{aligned} v_1 &= \sqrt{D} + \sqrt{D - \varepsilon\sqrt{D}}, \\ v_2 &= -\sqrt{D} - \sqrt{D + \varepsilon\sqrt{D}}, \\ v_3 &= -\sqrt{D} + \sqrt{D + \varepsilon\sqrt{D}}, \\ v_4 &= \sqrt{D} - \sqrt{D - \varepsilon\sqrt{D}}, \end{aligned} \tag{7.15}$$

$$D = c^2 + 1. \tag{7.16}$$

Доказательство. Предположим, что неприводимый многочлен пятой степени x^5+ax+b разрешим в радикалах. Тогда его резольвента

$$x^6 + 8ax^5 + 40a^2x^4 + 160a^3x^3 + 400a^4x^2 + (512a^5 - 3125b^4)x + (256a^6 - 9375ab^4)$$

имеет рациональные корни [Б11, а]. Следовательно, r удовлетворяет уравнению

$$(r + 2a)^4 (r^2 + 16a^2) - 5^5 b^4 (r + 3a) = 0, \tag{7.17}$$

откуда вытекает, что $r \neq -2a$, $r \neq -3a$ при $a \neq 0$. Определим неотрицательное рациональное число c и ненулевое рациональное число e, полагая

$$\varepsilon c = \frac{3r - 16a}{4(r + 3a)}, \quad e = \frac{-5b\varepsilon}{2(r + 2a)}, \tag{7.18}$$

где $\varepsilon = \pm 1$.

Тогда

$$c^2 + 1 = \frac{25(r^2 + 26a^2)}{16(r + 3a)^2},$$

$$3 - 4\varepsilon c = \frac{25a}{r + 3a},$$

$$11\varepsilon + 2c = \frac{25\varepsilon(r + 2a)}{2(r + 3a)},$$

так что соотношения

$$\frac{5e^4(3 - 4\varepsilon c)}{c^2 + 1} = \frac{5^5 ab^4 (r + 3a)}{(r + 2a)^4 (r^2 + 16a^2)} = a$$

и

$$-\frac{4e^5(11\varepsilon+2c)}{c^2+1}=\frac{5^5b^5(r+3a)}{(r+2a)^4(r^2+16a^2)}=b$$

задают требуемую параметризацию.

Покажем теперь, что неприводимое уравнение пятой степени

$$x^5+\frac{5e^4(3-4\varepsilon c)}{c^2+1}x-\frac{4e^5(11\varepsilon+2c)}{c^2+1}=0 \tag{7.19}$$

разрешимо в радикалах при $e=1$, причем корни его задаются формулами (7.11). Тогда для произвольного e преобразование $x\mapsto ex$ дает требуемые корни (7.13).

Из соотношений (7.15) вытекает, что

$$\begin{aligned} &v_1+v_4=2\sqrt{D}, \quad v_2+v_3=-2\sqrt{D},\\ &v_1v_4=\varepsilon\sqrt{D}, \qquad v_2v_3=-\varepsilon\sqrt{D}, \end{aligned} \tag{7.20}$$

следовательно,

$$\begin{cases} v_1+v_2+v_3+v_4=0,\\ v_1v_4+v_2v_3=0. \end{cases} \tag{7.21}$$

Далее, из (7.14) мы получаем

$$u_1^5=\frac{v_1^2v_3}{D^2}, \quad u_2^5=\frac{v_3^2v_4}{D^2}, \quad u_3^5=\frac{v_2^2v_1}{D^2}, \quad u_4^5=\frac{v_4^2v_2}{D^2}. \tag{7.22}$$

Стало быть,

$$u_1u_4=-\frac{\varepsilon}{\sqrt{D}}, \quad u_2u_3=\frac{\varepsilon}{\sqrt{D}}, \tag{7.23}$$

$$u_1u_2^2=\frac{v_3}{D}, \quad u_3^2u_4=\frac{v_2}{D}, \quad u_1^2u_3=\frac{v_1}{D}, \quad u_4^2u_2=\frac{v_4}{D}, \tag{7.24}$$

$$\begin{aligned} &u_1^3u_2=\frac{\varepsilon v_1v_3}{D\sqrt{D}}, \qquad u_2^3u_4=-\frac{\varepsilon v_3v_4}{D\sqrt{D}},\\ &u_3^3u_1=-\frac{\varepsilon v_1v_2}{D\sqrt{D}}, \qquad u_4^3u_3=\frac{\varepsilon v_2v_4}{D\sqrt{D}}, \end{aligned} \tag{7.25}$$

что с учетом соотношений (7.21) дает уравнения (7.7) и (7.8).

Из соотношений (7.15), (7.22)–(7.25) следует, что

$$\begin{aligned} 5(u_1^2u_4^2+u_2^2u_3^2-u_1^3u_2-u_2^3u_4-u_3^3u_1-u_4^3u_3-u_1u_2u_3u_4)=\\ =\frac{5(3-4\varepsilon\sqrt{D-1})}{D}=\frac{5(3-4\varepsilon c)}{c^2+1}, \end{aligned} \tag{7.26}$$

$$\begin{aligned} 5\big((u_1^3u_3u_4+u_2^3u_1u_3+u_3^3u_2u_4+u_4^3u_1u_2)-\\ -(u_1u_3^2u_4^2+u_2u_1^2u_3^2+u_3u_2^2u_4^2+u_4u_1^2u_2^2)\big)- \end{aligned} \tag{7.27}$$

$$-(u_1^5+u_2^5+u_3^5+u_4^5)=-\frac{44\varepsilon+8\sqrt{D-1}}{D}=-\frac{4(11\varepsilon+2c)}{c^2+1},$$

т. е. уравнения (7.9) и (7.10). Таким образом, уравнение

$$x^5+\frac{5(3-4\varepsilon c)}{c^2+1}x-\frac{4(11\varepsilon+2c)}{c^2+1}=0$$

разрешимо в радикалах и его корни задаются формулой (7.11). □

Замечание. В трактате Вебера [Б11, а] (см. также [Б22, Б32]) условие разрешимости формулируется несколько иначе: *уравнение* (7.1) *тогда и только тогда решается в радикалах, когда оно либо приводимо, либо его коэффициенты* a *и* b *имеют вид*

$$a=\frac{3125\lambda\mu^4}{(\lambda-1)^4(\lambda^2-6\lambda+25)},$$

$$b=\frac{3125\lambda\mu^5}{(\lambda-1)^4(\lambda^2-6\lambda+25)},$$

где λ *и* μ *— некоторые рациональные числа,* $\lambda\neq 1$.

Займемся теперь вычислением групп Галуа уравнений пятой степени (7.1), разрешимых в радикалах. Для этого нам понадобятся так называемые транзитивные группы подстановок. Группа G подстановок степени n называется *транзитивной*, если для любых двух номеров i, j, $1\leqslant i,j\leqslant n$, в группе G существует хотя бы одна подстановка, переводящая номер i в номер j. Значение транзитивных групп для теории Галуа объясняется следующим утверждением.

Теорема 2. *Группа Галуа неприводимого многочлена транзитивна.*

Доказательство. Достаточно заметить, что если многочлен неприводим, то все его корни $\alpha_1, \ldots, \alpha_n$ сопряжены между собой, и поэтому для любой пары корней α_i, α_j в поле $\mathbb{Q}(\alpha_1, \ldots, \alpha_n)$ найдется автоморфизм над $\mathbb{Q}$, переводящий корень α_i в корень α_j. □

Докажем теперь следующую общую теорему о транзитивных группах.

Теорема 3. *Порядок транзитивной группы подстановок степени n делится на n.*

Доказательство. Пусть G — произвольная транзитивная группа подстановок степени n. Разобьем группу G на непересекающиеся классы, объединяя в один класс все подстановки, одинаково действующие на номер 1. Так как группа G транзитивна, то число таких классов равно n. Обозначим через H класс подстановок, оставляющих номер n на месте. Очевидно, что H является подгруппой в группе G. Две подстановки $g_1, g_2 \in G$ принадлежат одному классу тогда и только тогда, когда $g_1 g_2^{-1} \in H$, т. е. когда подстановки g_1, g_2 принадлежат одному и тому же смежному классу по подгруппе H. Следовательно, рассматриваемые классы совпадают со смежными классами по подгруппе H. Поэтому индекс подгруппы H равен n. Так как группа G обладает подгруппой индекса n, то ее порядок делится на n. □

Простейшими транзитивными группами являются циклические группы. Очевидно, что циклическая группа подстановок степени n транзитивна тогда и только тогда, когда ее образующей служит цикл длины n. В частности, порядок такой группы равен n.

Нетрудно проверить, что число всех циклов длины n равно $(n-1)!$. В самом деле, любой цикл длины n единственным образом представляется в виде $(1 i_2 i_3 \ldots i_n)$, где $i_2, i_3, \ldots, i_n$ — некоторая перестановка чисел $2, 3, \ldots, n$. Так как циклическая группа порядка n содержит $\varphi(n)$ образующих, то отсюда следует, что число всех циклических транзитивных групп подстановок степени n равно $(n-1)!/\varphi(n)$. В частности, для $n = 5$ это число равно шести.

Очевидно, что группа подстановок, содержащая цикл длины n, транзитивна. Покажем, что в случае, когда n является простым числом, верно и обратное.

Теорема 4. *Любая транзитивная группа подстановок простой степени p содержит цикл длины p.*

Доказательство. Пусть G — произвольная транзитивная группа подстановок степени p. Разобьем множество всех циклов длины p, не принадлежащих группе G, на классы, объединяя циклы c_1 и c_2 в один класс, если в группе G найдется такой элемент g, что $c_2 = g^{-1} c_1 g$. Обозначим через H_c класс, содержащий цикл c. Легко видеть, что для любых двух циклов c_1 и c_2 длины p, не принадлежащих группе G, классы H_{c_1} и H_{c_2} либо совпадают, либо не пересекаются.

Рассмотрим теперь некоторый цикл $c = (i_1, \ldots, i_p)$ длины p, не принадлежащий группе G. Пусть g — произвольный элемент группы G. Если g имеет вид

$$g = \begin{pmatrix} i_1 & \ldots & i_p \\ j_1 & \ldots & j_p \end{pmatrix}, \tag{7.28}$$

то

$$g^{-1} c g = (j_1 j_2 \ldots j_p),$$

т. е. подстановка $g^{-1} c g$ также является циклом длины p. Кроме того, поскольку $g(g^{-1} c g) g^{-1} = c$, цикл $g^{-1} c g$ не принадлежит группе G. Следовательно, формула

$$\alpha(g) = g^{-1} c g$$

задает отображение α группы G на класс H_a.

Покажем, что отображение α является биекцией. Для этого достаточно установить, что из равенства $\alpha(g_1) = \alpha(g_2)$ следует равенство $g_1 = g_2$. Итак, пусть $\alpha(g_1) = \alpha(g_2)$, т. е.

$$g_1^{-1} c g_1 = g_2^{-1} c g_2.$$

Тогда

$$(g_1 g_2^{-1})^{-1} c (g_1 g_2^{-1}) = c,$$

т. е. $g^{-1} c g = c$, где $g = g_1 g_2^{-1}$. Поэтому, если подстановка g имеет вид (7.28), то $c = (j_1 j_2 \ldots j_p)$, откуда следует, что $g = c^k$, где k — такой номер, что $j_1 = j_{k+1}$. Число k взаимно просто с p, поскольку оно меньше p. Если $k > 0$, то найдутся такие числа a и b, что $ka + pb = 1$. Значит

$$c = c^{ka+pb} = c^{ka} = g^a,$$

и следовательно, вопреки предположению, $c \in G$. Поэтому $k = 0$, т. е. $g = e$ и $g_1 = g_2$.

Таким образом, все классы H_c состоят из одного и того же числа элементов, равного порядку группы G. Поэтому число всех циклов длины p, не принадлежащих группе G делится на порядок группы G и, следовательно, в силу теоремы 3, делится на p. С другой стороны, как было указано выше, число всех циклов длины p равно $(p-1)!$ и поэтому на p не делится. Следовательно, группа G обязательно содержит циклы длины p. □

Каждый цикл длины p, содержащийся в группе G, определяет циклическую подгруппу, состоящую из $p-1$ циклов длины p и тождественной подстановки. Так как эти циклические подгруппы пересекаются только по тождественной подстановке, то общее число циклов, содержащихся в группе G, равно $(p-1)m$, где m — число циклических подгрупп порядка p группы G. Следовательно, обозначая через ps число циклов длины p, не принадлежащих G, мы получим уравнение

$$(p-1)\,m + ps = (p-1)!,$$

откуда

$$m = (p-2)! - pr, \tag{7.29}$$

где $0 \leqslant r < \dfrac{(p-2)!}{p}$.

Для $p = 5$ число r может принимать лишь значения 0 и 1. В соответствии с этим мы получаем, что $m = 1$ или 6, т. е. транзитивная группа пятой степени содержит либо одну циклическую группу пятого порядка, либо шесть таких групп. Если в транзитивной группе G подстановок пятой степени содержится шесть циклических групп пятого порядка, то в силу тождеств

$$(ij)\,(kl) = (ikjlm)\,(ikjml),$$
$$(ij)\,(lk) = (ikjlm)\,(ikmlj),$$

группа G содержит произведения двух любых транспозиций и поэтому содержит любую четную подстановку. Следовательно, в этом случае G является либо знакопеременной группой A_5, либо симметрической группой S_5 и, значит, не разрешима.

Таким образом, нас будут интересовать лишь те группы G, которые содержат ровно одну циклическую группу пятого порядка. Пусть u — образующая этой группы. Для определенности

положим $u = (12345)$. Циклическую группу с этой образующей обозначим через $\mathbb{Z}_5$.

Рассмотрим, далее, подстановку $v = (25)\,(34)$. Непосредственно проверяется, что

$$e, u, u^2, u^3, u^4, v, uv, u^2v, u^3v, u^4v$$

образуют группу. Эта группа называется *диэдральной группой* порядка 5 и обозначается D_5. Она содержит циклическую группу $\mathbb{Z}_5$ в качестве нормального делителя, причем факторгруппа $D_5/\mathbb{Z}_5$ изоморфна группе $\mathbb{Z}_2$. Стало быть, группа D_5 разрешима.

Рассмотрим теперь подстановку $w = (2354)$. Легко видеть, что

$$w^4 = e, \quad wu = u^3w.$$

Поэтому подстановки

$$e, u, u^2, u^3, u^4,$$
$$w, uw, u^2w, u^3w, u^4w,$$
$$w^2, uw^2, u^2w^2, u^3w^2, u^4w^2,$$
$$w^3, uw^3, u^2w^3, u^3w^3, u^4w^3$$

образуют группу двадцатого порядка. Эта группа называется *метациклической группой* или же *группой Фробениуса* и обозначается F_{20}. Так как $w^2 = v$, то D_5 содержится в F_{20} в качестве нормального делителя индекса 2. Поэтому группа F_{20} разрешима.

Покажем, что группами $\mathbb{Z}_5$, D_5 и F_{20} исчерпываются все транзитивные группы пятой степени, содержащие только одну циклическую группу пятого порядка. В самом деле, пусть s — произвольная подстановка такой группы G. Если s переводит номер 1 в номер k, то подстановка $g = su^{-k}$ оставляет единицу на месте. Пусть g записывается в виде

$$g = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & i_2 & i_3 & i_4 & i_5 \end{pmatrix}.$$

Тогда

$$g^{-1}ug = (1i_2i_3i_4i_5).$$

Так как любой цикл длины 5 в группе G является по условию степенью цикла u, то отсюда следует, что

$$g^{-1}ug = u^k, \qquad k = 1, 2, 3, 4.$$

Поскольку $u = (12345)$, из этого соотношения вытекает, что подстановка g совпадает с одной из следующих подстановок

$$e = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 2 & 3 & 4 & 5 \end{pmatrix},$$

$$w = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 2 & 4 \end{pmatrix},$$

$$w^3 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix},$$

$$v = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 5 & 4 & 3 & 2 \end{pmatrix}$$

и поэтому принадлежит группе F_{20}. Следовательно, и подстановка s принадлежит этой группе. Значит, $G \subseteq F_{20}$. Но единственность подгруппами в F_{20}, содержащими группу $\mathbb{Z}_5$, являются $\mathbb{Z}_5$, D_5 и F_{20}, что и требовалось доказать.

Таким образом, разрешимое в радикалах уравнение пятой степени может иметь одну из трех групп Галуа: $\mathbb{Z}_5$, D_5, F_{20}. Выясним, в каких случаях разрешимое в радикалах уравнение (7.19) имеет в качестве группы Галуа ту или иную из этих групп. Для этого нам понадобится дискриминант многочлена пятой степени.

Нетрудно показать, что дискриминант Δ многочлена $x^5 + ax + b$ равен $\Delta = 4a^5 + 5^5 b^4$. В частности для многочлена, записанного в виде (7.19)

$$\Delta = \frac{4^4 5^5 e^{20}}{D^5}(4\varepsilon c^3 - 84c^2 - 37\varepsilon c - 122)^2. \qquad (7.30)$$

Воспользуемся теперь следующим фактом, доказательство которого можно найти, например, в замечательном учебнике по высшей алгебре Д. Граве [Б13]:

Теорема 5. *Многочлен x^5+ax+b имеет в точности один вещественный корень, если $\Delta > 0$.* □

В силу (7.30) отсюда следует, что уравнение (7.19) имеет ровно один вещественный корень. Но поскольку у уравнения (7.19) есть комплексные корни, $\mathbb{Z}_5$ его группой Галуа быть не может. Поэтому остаются две возможности: D_5 и F_{20}. Они также различаются дискриминантом Δ. А именно, можно доказать (см. [В5, В15]), что группа Галуа уравнения (7.19) изоморфна D_5 тогда и только тогда, когда рациональное число Δ является

полным квадратом. В виду (7.30) последнее условие эквивалентно тому, что полным квадратом является число $5D = 5,(c^2+1)$. Итак, мы получаем окончательный ответ: если $5D = 5(c^2+1)$ является квадратом рационального числа, то группа Галуа уравнения (7.19) есть D_5, в противном случае — группа F_{20}.

С помощью теоремы 1 нетрудно получить примеры уравнений пятой степени вида (7.1), не разрешимые в радикалах.

Пусть p — простое число, $p \equiv 3 \pmod 4$. Покажем, что уравнение пятой степени

$$x^5 + 2px + 2p^2 = 0$$

не разрешимо в радикалах. Заметим прежде всего, что в силу критерия Эйзенштейна многочлен $x^5 + 2px + 2p^2$ неприводим. Предположим, что наше уравнение разрешимо в радикалах. Тогда на основании теоремы 1 существуют такие рациональные числа $\varepsilon = \pm 1$, $c \geqslant 0$ и $e \neq 0$, что

$$2p = \frac{5e^4}{c^2+1}(3 - 4\varepsilon c), \qquad (7.31)$$

$$2p^2 = -\frac{4e^5}{c^2+1}(11\varepsilon + 2c). \qquad (7.32)$$

Представим рациональные числа c и e в виде несократимых дробей $c = m/n$ и $e = r/s$, где m, n и r, s — попарно взаимно простые целые числа. Подставляя эти выражения для c и e в формулы (7.31) и (7.32), получим

$$2p(m^2+n^2)s^4 = 5r^4(3n - 4\varepsilon m)n, \qquad (7.33)$$

$$2p^2(m^2+n^2)s^5 = -4r^5(11\varepsilon n + 2m)n. \qquad (7.34)$$

Так как p — простое число, $p \equiv 3 \pmod 4$ и $(m,n) = 1$, то m^2+n^2 не делится на p. Более того, поскольку $(r,s) = 1$, из соотношения (7.33) вытекает, что r также не делится на p. Пусть p^α, p^β, p^γ, p^δ — наибольшие степени числа p, делящие соответственно числа n, $3n - 4\varepsilon m$, $11\varepsilon n + 2m$ и s. Приравнивая степени числа p в обеих частях соотношений (7.33) и (7.34), получаем

$$\begin{cases} 1 + 4\delta = \alpha + \beta \\ 2 + 5\delta = \alpha + \gamma. \end{cases} \qquad (7.35)$$

С другой стороны, число p не может одновременно делить n и $3n - 4\varepsilon m$, поэтому либо $\alpha = 0$, либо $\beta = 0$. Аналогично, либо $\alpha = 0$, либо $\gamma = 0$ и либо $\beta = 0$, либо $\gamma = 0$. Другими словами, по меньшей мере два из трех чисел α, β, γ должны быть равными нулю, что противоречит системе (7.35). Следовательно, уравнение

$$x^2 + 2px + 2p^2 = 0$$

не разрешимо в радикалах.

ГЛАВА 7

РЕШЕНИЕ УРАВНЕНИЯ 5-Й СТЕПЕНИ

§ 1. Определение тэта-функций

Напомним, что функция f называется *двоякопериодической*, если для любых целых m и n

$$f(z + n\omega_1 + m\omega_2) = f(z),$$

причем $\omega_1/\omega_2 \notin \mathbb{R}$. Теорема 1 из § 2 главы 2 утверждает, что целая функция не может быть двоякопериодической.

Тэта-функции представляют собой целые функции, имеющие один настоящий период и один квазипериод: при добавлении к аргументу квазипериода функция хотя и изменяется, но по достаточно простому закону.

Верхнюю полуплоскость

$$H = \{\tau \in \mathbb{C} \mid \operatorname{Im}\tau > 0\}$$

можно отобразить внутрь единичного круга

$$D = \{q \in \mathbb{C} \mid |q| \leqslant 1\},$$

положив $q = e^{i\pi\tau}$. В самом деле, пусть $\tau = x + iy$, где $x, y \in \mathbb{R}$. Тогда $q = e^{i\pi x - \pi y} = e^{-\pi y}e^{i\pi x}$ и $|q| = e^{-\pi y}$. Поэтому $|q| < 1$ тогда и только тогда, когда $y > 0$, т. е. $\operatorname{Im}\tau > 0$.

Фиксируем число $\tau \in H$ и рассмотрим ряд

$$\theta_3(v \mid \tau) = \sum_{m=-\infty}^{\infty} e^{(m^2\tau + 2mv)\,\pi i} = \sum_{m=-\infty}^{\infty} q^{m^2} e^{2\pi i m v}$$

Модуль отношения последовательных членов этого ряда равен $|q^{2m+1}e^{2\pi i v}| \leqslant |q|^{2m+1}e^{2\pi|v|}$. А так как $\lim\limits_{m\to\infty} |q|^{2m+1} = 0$, то

$\theta_3(v\,|\,\tau)$ представляет собой ряд целых функций переменной v, равномерно сходящийся в любой области $|v| \leqslant c$, где c — константа. Поэтому $\theta_3(v\,|\,\tau)$ — целая функция переменной v. Эту функцию мы часто будем обозначать для краткости $\theta_3(v)$.

Наряду с функцией $\theta_3(v)$ удобно рассматривать функции

$$\theta_0(v) = \theta_3(v+1/2) = \sum q^{m^2} e^{2\pi i m v} e^{\pi i m} = \sum (-1)^m q^{m^2} e^{2\pi i m v},$$

$$\theta_1(v) = ie^{-\pi i(v-\tau/4)}\,\theta_3(v+1/2-\tau/2) =$$

$$= ie^{-\pi i(v-\tau/4)} \sum q^{m^2} e^{2\pi i m v} e^{\pi i m} e^{-\pi i m \tau} =$$

$$= i \sum (-1)^m q^{(m-1/2)^2} e^{\pi i (2m-1) v},$$

$$\theta_2(v) = e^{-\pi i(v-\tau/4)}\,\theta_3(v-\tau/2) =$$

$$= e^{-\pi i(v-\tau/4)} \sum q^{m^2} e^{2\pi i m v} e^{-\pi i m v} = \sum q^{(m-1/2)^2} e^{\pi i (2m-1) v}.$$

Функции $\theta_0(v|\tau)$, $\theta_1(v|\tau)$, $\theta_2(v|\tau)$, $\theta_3(v|\tau)$ называются *тэта-функциями.*

Теорема 1. *Для функций $\theta_k(v)$, $k = 0, 1, 2, 3$ имеют место соотношения*

$$\theta_k(v+1) = \theta_k(v), \qquad k = 0, 3;$$

$$\theta_k(v+1) = -\theta_k(v), \qquad k = 1, 2;$$

$$\theta_k(v+\tau) = A(v)\,\theta_k(v), \qquad k = 2, 3;$$

$$\theta_k(v+\tau) = -A(v)\,\theta_k(v), \qquad k = 0, 1,$$

где $A(v) = q^{-1}e^{-2\pi i v}$.

Доказательство. Проверим эти соотношения для функции θ_3. При замене v на $v+1$ ряд (1.1) не изменится, так как $e^{2\pi i m(v+1)} = e^{2\pi i m v}$. Это означает, что $\theta_3(v+1) = \theta_3(v)$. Далее

$$\theta_3(v+\tau) = \sum q^{m^2} e^{2\pi i m v} q^{2m} =$$

$$= q^{-1}e^{-2\pi i v} \sum q^{(m+1)^2} e^{2\pi i (m+1) v} = q^{-1}e^{-2\pi i v}\theta_3(v).$$

Оставшиеся утверждения проверяются аналогично. □

§ 2. Нули тэта-функций

Пусть числа m и k связаны соотношением $k = 1 - m$. Тогда $(-1)^m = -(-1)^k$ и $q^{(m-1/2)^2} = q^{(k-1/2)^2}$, а значит, $\theta_1(-v) = -\theta_1(v)$. В частности, $\theta_1(0) = 0$.

Так как $\theta_1(v+1) = -\theta_1(v)$ и $\theta_1(v+\tau) = -A(v)\,\theta_1(v)$, то $\theta_1(m+n\tau) = 0$. Покажем, что других нулей у функции θ_1 нет. Для этого достаточно проверить, что внутри параллелограмма П с вершинами $\pm 1/2 \pm \tau/2$ функция θ_1 имеет единственный нуль.

Ясно, что

$$\frac{\theta_1'(v+1)}{\theta_1(v+1)} = \frac{\theta_1'(v)}{\theta_1(v)},$$

$$\frac{\theta_1'(v+\tau)}{\theta_1(v+\tau)} = \frac{-A'(v)\,\theta_1(v) - A(v)\,\theta_1'(v)}{-A(v)\,\theta_1(v)} =$$

$$= \frac{A'(v)}{A(v)} + \frac{\theta_1'(v)}{\theta_1(v)} = -2\pi i + \frac{\theta_1'(v)}{\theta_1(v)}.$$

Следовательно, если C — граница параллелограмма П, обходимая против часовой стрелки, то

$$\frac{1}{2\pi i} \int\limits_C \frac{\theta_1'(z)}{\theta_1(z)}\,dz = 1$$

В самом деле, интегралы по сторонам, полученным сдвигом на 1, сокращаются, а сумма интегралов по сторонам, полученным сдвигом на τ, равна $2\pi i$, поскольку длины этих сторон равны 1.

Если $f(z) = c_k(z-a_k)^k + \ldots$, то

$$\frac{f'(z)}{f(z)} = \frac{k}{z-a_k} + \ldots$$

Поэтому для целой аналитической функции f величина

$$\frac{1}{2\pi i} \int\limits_C \frac{f'(z)}{f(z)}\,dz$$

равна количеству нулей с учетом их кратности, лежащих внутри C. Таким образом, внутри параллелограмма П функция θ_1 имеет ровно один нуль.

Исходя из выражений для функций θ_0, θ_1 и θ_2 через функцию θ_3, легко получить следующую таблицу нулей тэта-функций:

	$\theta_0(v)$	$\theta_1(v)$	$\theta_2(v)$	$\theta_3(v)$
нули	$m+(n+1/2)\tau$	$m+n\tau$	$m+1/2+n\tau$	$m+1/2+(n+1/2)\tau$

Задачи.

1. Доказать, что функции θ_0, θ_2 и θ_3 четные.

§ 3. Соотношение $\theta_3^4=\theta_2^4+\theta_0^4$

Величины $\theta_i=\theta_i(0)$, $i=0,2,3$, и $\theta_1'=\theta_1'(0)$ называют тэта-константами. Напомним, что они зависят от параметра τ.

Рассмотрим функцию $f(v)=\dfrac{a\theta_2^2(v)+b\theta_3^2(v)}{\theta_0^2(v)}$, где $a,b\in\mathbb{C}$. Числа 1 и τ являются периодами функции f, поэтому f — двоякопериодическая функция с фундаментальным параллелограммом П. Фундаментальный параллелограмм можно сдвинуть так, чтобы внутри него лежал лишь один нуль функции $\theta_0(v)$, а именно точка $\tau/2$. Если числа a и b подобраны так, что

$$a\theta_2^2(\tau/2)+b\theta_3^2(\tau/2)=0,$$

то эллиптическая функция $f(v)$ имеет внутри параллелограмма П не более чем однократный полюс. Следовательно, она постоянна.

Подставив значения $v=\tau/2$ и $v=0$ в соотношение

$$\theta_2(v)=e^{-\pi i(v-\tau/4)}\theta_3(v-\tau/2),$$

получим соответственно

$$\theta_2(\tau/2)=e^{-\pi i\tau/4}\theta_3(0),$$

$$\theta_2(0)=e^{\pi i\tau/4}\theta_3(-\tau/2)=e^{\pi i\tau/4}\theta_3(\tau/2).$$

Поэтому $a\theta_2^2(\tau/2)+b\theta_3^2(\tau/2)=aB^2\theta_3^2+bB^2\theta_2^2$, где $B=e^{-\pi i\tau/4}$.

Положим $a=-\theta_2^2$ и $b=\theta_3^2$. Тогда $aB^2\theta_3^2+bB^2\theta_2^2=0$, а значит,

$$-\theta_2^2(v)\theta_2^2+\theta_3^2(v)\theta_3^2=c\theta_0^2(v).$$

Чтобы вычислить константу c, положим $v=1/2$. Функция θ_2 в этой точке обращается в нуль, а для вычисления $\theta_3(1/2)$ и $\theta_0(1/2)$ можно подставить значения $v=0$ и $v=1/2$ в соотношение $\theta_0(v)=\theta_3(v+1/2)$. В результате получим $\theta_0=\theta_3(1/2)$ и $\theta_0(1/2)=\theta_3(1)=\theta_3$. Поэтому

$$\theta_0^2\theta_3^2=c\theta_3^2,$$

т. е. $c=\theta_0^2$. Таким образом, $\theta_3^2(v)\theta_3^2-\theta_2^2(v)\theta_2^2=\theta_0^2(v)\theta_0^2$. В частности, при $v=0$ получаем

$$\theta_3^4=\theta_2^4+\theta_0^4,$$

т. е. $(1+2q+2q^4+2q^9+\ldots)^4=16q(1+q^{1\cdot2}+q^{2\cdot3}+q^{3\cdot4}+\ldots)^4+(1-2q+2q^4-2q^9+\ldots)^4$.

Задачи.

1. Докажите следующие соотношения:

$$\theta_2^2(v)\theta_0^2=\theta_0^2(v)\theta_2^2-\theta_1^2(v)\theta_3^2;$$
$$\theta_3^2(v)\theta_0^2=\theta_0^2(v)\theta_3^2-\theta_1^2(v)\theta_2^2;$$
$$\theta_1^2(v)\theta_0^2=\theta_3^2(v)\theta_2^2-\theta_2^2(v)\theta_3^2.$$

§ 4. Представление тэта-функций бесконечными произведениями

Функция $\theta_3(v)=\sum\limits_{k=-\infty}^{\infty}q^{k^2}e^{2\pi ikv}$ имеет нули

$$m+1/2+(n+1/2)\tau.$$

Положим $s=e^{2\pi iv}$. При этом преобразовании нули функции $\theta_3(v)$ переходят в точки $e^{2\pi i(m+1/2)}e^{2\pi i(n+1/2)\tau}=-q^{2n+1}$, где $q=e^{\pi i\tau}$.

Разобьем эти точки на два множества:

$$-q^{-1}, -q^{-3}, -q^{-5}, \dots \tag{4.1}$$

$$-q, -q^3, -q^5, \dots \tag{4.2}$$

Предельной точкой множества (4.1) является ∞, а предельной точкой множества (4.2) является 0.

Ряд $\sum\limits_{k=1}^{\infty} |q^{2k-1}|$ сходится, поэтому функция

$$f_1(s) = \prod_{k=1}^{\infty} (1 + q^{2k-1}s)$$

будет целой функцией от s с нулями в точках множества (4.1). Аналогично функция

$$f_2(s) = \prod_{k=1}^{\infty} (1 + q^{2k-1}s^{-1})$$

будет целой функцией, если ее рассматривать как функцию от s^{-1} (как функция от s она имеет особенность в нуле). Нулями функции $f_2(s)$ являются точки множества (4.2).

Рассмотрим функцию $f(s) = f_1(s) f_2(s)$. Функция

$$g(v) = f(e^{2\pi i v}) = f(s)$$

имеет те же нули, что и функция $\theta_3(v)$. При замене $v \mapsto v + 1$ величина $s = e^{2\pi i v}$ не изменяется, поэтому $g(v+1) = g(v)$. При замене $v \mapsto v + \tau$ величина s заменяется на $e^{2\pi i \tau} e^{2\pi i v} = sq^2$, поэтому

$$g(v+\tau) = \frac{1 + q^{-1}s^{-1}}{1 + qs}\, g(v) = q^{-1}e^{-2\pi i v} g(v).$$

Следовательно, отношение функций $\theta_3(v)$ и $g(v)$ является целой двоякопериодической функцией, т. е. константой. Таким образом,

$$\theta_3(v) = c \prod_{k=1}^{\infty} (1 + q^{2k-1}e^{2\pi i v})(1 + q^{2k-1}e^{-2\pi i v}) \tag{4.3}$$

Аналогичные разложения можно получить и для остальных тэта-функций. Так как $e^{2\pi i(v+1/2)} = e^{-2\pi i v}$, то

$$\theta_0(v) = \theta_3(v + 1/2) = c \prod_{k=1}^{\infty} (1 - q^{2k-1}e^{2\pi i v})(1 - q^{2k-1}e^{-2\pi i v}) \tag{4.4}$$

Так как

$$q^{2k-1}e^{2\pi i(v+1/2-\tau/2)} = -q^{2k-2}e^{2\pi i v},$$

$$q^{2k-1}e^{-2\pi i(v+1/2-\tau/2)} = -q^{2k}e^{-2\pi i v},$$

то

$$\theta_1(v) = ie^{-\pi i v} q^{1/4} \theta_3(v + 1/2 - \tau/2) =$$

$$ci(1 - e^{2\pi i v})\, e^{-\pi i v} q^{1/4} \prod_{k=1}^{\infty} (1 - q^{2k}e^{2\pi i v})(1 - q^{2k}e^{-2\pi i v}).$$

Ясно также, что $i(1 - e^{2\pi i v})\, e^{-\pi i v} = 2 \sin \pi v$, поэтому

$$\theta_1(v) = 2cq^{1/4} \sin \pi v \prod_{k=1}^{\infty} (1 - q^{2k}e^{2\pi i v})(1 - q^{2k}e^{-2\pi i v}). \tag{4.5}$$

Легко проверить, что $\theta_2(v) = \theta_1(v + 1/2)$, поэтому

$$\theta_2(v) = 2cq^{1/4} \cos \pi v \prod_{k=1}^{\infty} (1 + q^{2k}e^{2\pi i v})(1 + q^{2k}e^{-2\pi i v}). \tag{4.6}$$

Докажем теперь, что

$$c = \prod_{k=1}^{\infty} (1 - q^{2k}). \tag{4.7}$$

Рассмотрим для этого последовательность функций

$$F_n(s) = \prod_{k=1}^{n} (1 - q^{2k-1}s)(1 - q^{2k-1}s^{-1}) = \sum_{k=-n}^{n} a_k(n)\, s^k.$$

Она равномерно сходятся к функции

$$\prod_{k=1}^{\infty} (1 - q^{2k-1}s)(1 - q^{2k-1}s^{-1}) = \frac{1}{c}\theta_0(v) = \frac{1}{c} \sum_{k=-\infty}^{\infty} (-1)^k q^{k^2} s^k.$$

Приравнивая коэффициенты при нулевой степени s, получим $1/c = \lim\limits_{n\to\infty} a_0(n)$.

Ясно, что $a_n(n) = (-1)^n q^{1+3+\dots+(2n-1)} = (-1)^n q^{n^2}$. Кроме

того, очевидно равенство

$$\frac{f_n(q^2s)}{f_n(s)} = \frac{(1-q^{2n+1}s)(1-q^{-1}s^{-1})}{(1-qs)(1-q^{2n-1}s^{-1})} = -\frac{1-q^{2n+1}s}{qs-q^{2n}},$$

поэтому

$$(qs-q^{2n})\sum_{k=-n}^{n} a_k(n)\,q^{2k}s^k = -(1-q^{2n+1}s)\sum_{k=-n}^{n} a_k(n)\,s^k,$$

т. е.

$$\sum_{k=-n}^{n} a_k(n)\,(1-q^{2(n+k)})\,s^k = \sum_{k=-n}^{n} a_k(n)\,(q^{2n+1}-q^{2k+1})\,s^{k+1}.$$

Следовательно,

$$a_0(n) = q^{n^2}\frac{\prod_{k=1}^{n}(1-q^{2(n+k)})}{\prod_{k=1}^{n}(q^{2k+1}-q^{2n+1})} = \frac{\prod_{k=1}^{n}(1-q^{2(n+k)})}{\prod_{k=1}^{n}(1-q^{2k})}.$$

Пусть $|q|^2 = \alpha < 1$. Тогда $1-\alpha^n \leqslant |1-q^{2(n+k)}| \leqslant 1+\alpha^n$. Кроме того, $\lim\limits_{n\to\infty} n\ln(1\pm\alpha^n) = 0$. Поэтому $\lim\limits_{n\to\infty}\prod\limits_{k=1}^{n}(1-q^{2(n+k)}) = 1$, а значит,

$$c = \lim_{n\to\infty}\frac{1}{a_0(n)} = \prod_{k=1}^{\infty}(1-q^{2k}).$$

§ 5. Соотношение $\theta_1'(0) = \pi\theta_0(0)\theta_2(0)\theta_3(0)$

Из формул (4.4)–(4.7) следует, что

$$\theta_0(0) = c\prod_{k=1}^{\infty}(1-q^{2k-1})^2; \tag{5.1}$$

$$\theta_2(0) = 2q^{1/4}c\prod_{k=1}^{\infty}(1+q^{2k})^2; \tag{5.2}$$

$$\theta_3(0) = c\prod_{k=1}^{\infty}(1+q^{2k-1})^2; \tag{5.3}$$

$$\theta_1'(0) = 2\pi q^{1/4}c\prod_{k=1}^{\infty}(1-q^{2k})^2. \tag{5.4}$$

Для доказательства последней формулы достаточно заметить, что $\theta_1'(0) = \lim\limits_{v\to\infty}\theta_1(v)/v$.

Так как $\prod\limits_{k=1}^{\infty}(1-q^{2k})^2 = c^2$, то $\theta_1'(0) = 2\pi q^{1/4}c^3$. Поэтому для доказательства соотношения $\theta_1'(0) = \pi\theta_0(0)\theta_2(0)\theta_3(0)$ достаточно проверить, что

$$\prod_{k=1}^{\infty}(1-q^{2k-1})(1+q^{2k})(1+q^{2k-1}) = 1.$$

Ясно, что

$$\prod_{k=1}^{\infty}(1-q^{2k-1}) = \prod_{n=1}^{\infty}(1-q^n)(1-q^{2n})^{-1}$$

и

$$\prod_{k=1}^{\infty}(1+q^{2k})(1+q^{2k-1}) = \prod_{n=1}^{\infty}(1+q^n).$$

Поэтому

$$\prod_{k=1}^{\infty}(1-q^{2k-1})(1+q^{2k})(1+q^{2k-1}) =$$

$$= \prod_{n=1}^{\infty}(1-q^n)(1+q^n)(1-q^{2n})^{-1} = \prod_{n=1}^{\infty}(1-q^{2n})(1-q^{2n})^{-1} = 1.$$

§ 6. η-функция Дедекинда и функции f, f_1, f_2

Формулы (5.1)–(5.4) с учетом (4.7) позволяют представить тэта-константы в следующем виде:

$$\theta_1'(0) = 2\pi\eta^3(\tau), \quad \text{где} \quad \eta(\tau) = q^{1/12}\prod_{k=1}^{\infty}(1-q^{2k});$$

$$\theta_3(0) = \eta(\tau)f^2(\tau), \quad \text{где} \quad f(\tau) = q^{-1/24}\prod_{k=1}^{\infty}(1+q^{2k-1});$$

$$\theta_0(0) = \eta(\tau) f_1^2(\tau), \quad \text{где} \quad f_1(\tau) = q^{-1/24} \prod_{k=1}^{\infty} (1 - q^{2k-1});$$

$$\theta_2(0) = \eta(\tau) f_2^2(\tau), \quad \text{где} \quad f_2(\tau) = \sqrt{2} q^{1/12} \prod_{k=1}^{\infty} (1 + q^{2k}).$$

Из соотношения $\theta_3^4(0) = \theta_2^4(0) + \theta_0^4(0)$ следует, что

$$f^8 = f_1^8 + f_2^8. \tag{6.1}$$

В предыдущем параграфе было показано, что

$$\prod_{k=1}^{\infty} (1 - q^{2k-1})(1 + q^{2k})(1 + q^{2k-1}) = 1.$$

Поэтому

$$f f_1 f_2 = \sqrt{2}.$$

Функции f, f_1 и f_2 можно следующим образом выразить через функцию η:

$$f(\tau) = \frac{e^{-\pi i/24} \eta(\tau/2 + 1/2)}{\eta(\tau)}; \tag{6.3}$$

$$f_1(\tau) = \frac{\eta(\tau/2)}{\eta(\tau)}; \tag{6.4}$$

$$f_2(\tau) = \sqrt{2} \frac{\eta(2\tau)}{\eta(\tau)}. \tag{6.5}$$

Докажем, например, формулу (6.3). Легко проверить, что

$$\frac{\eta(\tau/2 + 1/2)}{\eta(\tau)} = q^{-1/24} e^{\pi i/24} \prod_{k=1}^{\infty} \frac{1 - (iq^{1/2})^{2k}}{1 - q^{2k}} =$$

$$= q^{-1/24} e^{\pi i/24} \frac{(1+q)(1-q^2)(1+q^3)(1-q^4)\ldots}{(1-q^2)(1-q^4)\ldots} =$$

$$= q^{-1/24} e^{\pi i/24} \prod_{k=1}^{\infty} (1 - q^{2k-1}).$$

Формулы (6.4) и (6.5) доказываются еще проще.

§ 7. Преобразования тэта-функций по параметру τ

В § 1 мы выяснили, как ведут себя функции $\theta_i(v \mid \tau)$ при заменах v на $v+1$ и на $v+\tau$. Оказывается, что при заменах τ на $\tau+1$ и на $-1/\tau$ функции $\theta_i(v \mid \tau)$ тоже преобразуются по достаточно простым законам. Для замены τ на $\tau+1$ это не удивительно, поскольку при такой замене величина $q = e^{\pi i \tau}$ заменяется на $q' = -q$. Поэтому

$$\theta_0(v \mid \tau + 1) = \theta_3(v \mid \tau), \quad \theta_3(v \mid \tau + 1) = \theta_0(v \mid \tau) \quad \text{и}$$

$$\theta_k(v \mid \tau + 1) = e^{\pi i/4} \theta_k(v \mid \tau) \quad \text{при} \quad k = 1, 2.$$

Что же касается замены τ на $\tau' = -1/\tau$, то при этом величина $q = e^{-\pi i \tau}$ заменяется на $q' = e^{-\pi i/\tau} = e^{-\pi^2/\ln q}$. Наличие закона преобразования при такой замене, безусловно, удивительно.

Чтобы найти этот закон, рассмотрим функцию

$$g(v) = e^{\pi i \tau' v^2} \frac{\theta_3(\tau' v \mid \tau')}{\theta_3(v \mid \tau)}.$$

Несложные вычисления показывают, что $g(v+1) = g(v)$ и $g(v+\tau) = g(v)$, т. е. g — двоякопериодическая функция. При этом нули знаменателя имеют вид

$$v = (m + 1/2)\tau + (n + 1/2),$$

а нули числителя определяются соотношением

$$\tau' v = (m + 1/2)\tau' + (n + 1/2),$$

т. е.

$$v = (m + 1/2) - (n + 1/2)\tau.$$

Таким образом, нули числителя совпадают с нулями знаменателя, а значит, g — целая двоякопериодическая функция. Следовательно, $g = C$, где C — константа.

Заменяя v на $v + 1/2$, $v + \tau/2$ и $v + 1/2 + \tau/2$, из равенства

$$\theta_3(\tau' v \mid \tau') = C e^{-\pi i \tau' v^2} \theta_3(v \mid \tau) \tag{7.1}$$

можно получить равенства

$$\theta_2(\tau' v \mid \tau') = C\, e^{-\pi i \tau' v^2} \theta_0(v \mid \tau), \tag{7.2}$$

$$\theta_0(\tau' v \mid \tau') = C\, e^{-\pi i \tau' v^2} \theta_2(v \mid \tau), \tag{7.3}$$

$$\theta_1(\tau' v \mid \tau') = iC\, e^{-\pi i \tau' v^2} \theta_1(v \mid \tau). \tag{7.4}$$

Перемножив равенства (7.1)–(7.3) и положив $v = 0$, получим

$$\theta_2(0 \mid \tau')\,\theta_3(0 \mid \tau')\,\theta_0(0 \mid \tau') = C^3\,\theta_2(0 \mid \tau)\,\theta_3(0 \mid \tau)\,\theta_0(0 \mid \tau). \tag{7.5}$$

Продифференцировав по v равенство (7.4) и положив $v = 0$, получим

$$\tau'\,\theta_1'(0 \mid \tau') = iC\,\theta_1'(0 \mid \tau). \tag{7.6}$$

А так как $\theta_1' = \pi\theta_0\theta_2\theta_3$, то из равенств (7.5) и (7.6) следует, что $C^2 = -i\tau$. Ясно так же, что при чисто мнимом τ обе величины $\theta_3(0 \mid \tau)$ и $\theta_3(0 \mid \tau')$ положительны. Следовательно, $C = \sqrt{-i\tau}$ и

$$\theta_3(0 \mid -1/\tau) = \sqrt{-i\tau}\,\theta_3(0 \mid \tau). \tag{7.7}$$

§ 8. Преобразования η-функции Дедекинда

Преобразования тэта-функций по параметру τ приводит к преобразованиям η-функции Дедекинда, так как

$$2\pi\eta^3(\tau) = \theta_1'(0 \mid \tau).$$

Из соотношения $\theta_1(0 \mid \tau + 1) = e^{\pi i/4}\theta_1(0 \mid \tau)$ получаем

$$\eta(\tau + 1) = e^{\pi i/12}\eta(\tau). \tag{8.1}$$

Если учесть, что $C = \sqrt{-i\tau}$ и $\tau' = -1/\tau$, то соотношение (7.6) принимает вид

$$\theta_1'(0 \mid -1/\tau) = (\sqrt{-i\tau})^3\,\theta_1'(0 \mid \tau).$$

Следовательно,

$$\eta(-1/\tau) = \sqrt{-i\tau}\,\eta(\tau). \tag{8.2}$$

С помощью формул (8.1) и (8.2) можно получить законы преобразования функций f, f_1, и f_2 при заменах τ на $\tau + 1$ и на $-1/\tau$. Например,

$$f(\tau+1) = \frac{e^{-\pi i/24}\eta(\tau/2+1)}{\eta(\tau+1)} = \frac{e^{-\pi i/24}\eta\left(\tau/2\right)}{\eta(\tau)} = e^{-\pi i/24} f_1(\tau). \tag{8.3}$$

Аналогично получаем

$$f_1(\tau + 1) = e^{-\pi i/24} f(\tau), \tag{8.4}$$

$$f_2(\tau + 1) = e^{\pi i/12} f_2(\tau). \tag{8.5}$$

Отметим, что $f(\tau + 2) = e^{-\pi i/12} f(\tau)$, а значит, $f(\tau + 48) = f(\tau)$.

Легко проверить, что

$$f_1\left(-1/\tau\right) = \frac{\eta\left(\dfrac{1}{2\tau}\right)}{\eta\left(-1/\tau\right)} = \frac{\sqrt{-2i\tau}\,\eta(2\tau)}{\sqrt{-i\tau}\,\eta(\tau)} = f_2(\tau). \tag{8.6}$$

Заменив τ на $-1/\tau$, получим

$$f_2\left(-1/\tau\right) = f_1(\tau). \tag{8.7}$$

Выражение для $f(-1/\tau)$ аналогичными вычислениями получить не удается. Но если воспользоваться тем, что

$$f(\tau)\, f_1(\tau)\, f_2(\tau) = \sqrt{2}$$

и

$$f\left(-1/\tau\right) f_1(\tau)\, f_2(\tau) = f\left(-1/\tau\right) f_2\left(-1/\tau\right) f_1\left(-1/\tau\right) = \sqrt{2},$$

то получим

$$f\left(-1/\tau\right) = f(\tau). \tag{8.8}$$

Докажем теперь соотношение

$$f(\tau)\, f\left(\frac{\tau - 1}{\tau + 1}\right) = \sqrt{2}. \tag{8.9}$$

Из (6.4) и (6.5) следует, что $f_1(2\tau)\, f_2(\tau) = \sqrt{2}$. А так как

$$f_1(2\tau) = e^{-\pi i/24} f(2\tau - 1)$$

и

$$f_2(\tau) = f_1(-1/\tau) = e^{\pi i/24} f(1 - 1/\tau),$$

то

$$f(2\tau - 1)\, f(1 - 1/\tau) = \sqrt{2}.$$

Положим $x = 2\tau - 1$. Тогда $1 - 1/\tau = \dfrac{x-1}{x+1}$, а значит, полученное соотношение эквивалентно (8.9).

§ 9. Общая схема решения уравнения пятой степени

Оставшаяся часть этой главы непосредственно посвящена решению уравнения пятой степени. Соответствующая конструкция достаточно сложна и опирается на разнообразные факты, причем доказательства многих из них требуют громоздких вычислений. Чтобы помочь читателю разобраться в этой конструкции, мы сейчас опишем общую схему действий: что и с какой целью делается.

Пусть $f(\tau)$, $f_1(\tau)$, $f_2(\tau)$ — определенные в § 6 функции. Положим

$$u = f(\tau), \quad v_c = f\left(\frac{\tau + c}{5}\right), \quad v_\infty = f(5\tau).$$

Исследование поведения u и v, где символом v обозначена одна из функций v_c или v_∞, при заменах τ на $\tau + 2$, $-\dfrac{1}{\tau}$ и $\dfrac{\tau - 1}{\tau + 1}$ показывает, что uv и u/v при этих заменах преобразуются следующим образом:

	uv	$\dfrac{u}{v}$
$\tau + 2$	$e^{-\pi i/2} uv$	$e^{\pi i/3} \dfrac{u}{v}$
$-1/\tau$	uv	$\dfrac{u}{v}$
$\dfrac{\tau - 1}{\tau + 1}$	$-\dfrac{2}{uv}$	$-\dfrac{v}{u}$

При этом индекс c функции v_c преобразуется точно так же, как преобразуется τ. Эти факты будут доказаны в §§ 10–13.

Используя эти наблюдения несложно доказать, что

$$\left(\frac{u}{v}\right)^3 + \left(\frac{v}{u}\right)^3 = (uv)^2 - \frac{4}{(uv)^2},$$

т. е.

$$au^6 - u^5 v^6 + 4uv + u^6 = 0.$$

Будем рассматривать $u = f(\tau)$ как параметр. Для каждого значения этого параметра получаем уравнение шестой степени, корни которого явным образом выражаются через τ. Нам бы хотелось сделать то же самое для уравнения $y^5 + 5y = a$. Оказывается, что этого можно достичь следующим образом. Рассмотрим многочлен пятой степени с корнями

$$w_z = \frac{(v_\infty - v_z)(v_{z+1} - v_{z-1})(v_{z+2} - v_{z-2})}{\sqrt{5}\, u^3}, \tag{9.1}$$

где $z = 0, 1, 2, 3, 4$. Вычисления показывают, что этот многочлен имеет вид

$$w(w^2 + 5)^2 - u^{12} + 64u^{-12}.$$

После замены

$$y(\tau) = \frac{f_1^8(\tau) - f_2^8(\tau)}{f^2(\tau)\,(w^2(\tau) + 5)} \tag{9.2}$$

мы приходим к уравнению

$$y^5 + 5y = \frac{f_1^8 - f_2^8}{f^2}.$$

Итак, для решения уравнения $y^5 + 5y = a$ нужно поступить следующим образом. Сначала находим τ, для которого

$$f_1^8(\tau) - f_2^8(\tau) = af^2(\tau).$$

Эта задача сводится к решению квадратного уравнения и вычислению обратной функции для функции f. Затем по формуле (9.1) вычисляем $w_z(\tau)$, а после этого по формуле (9.2) вычисляем $y_z(\tau)$. Это и есть корни рассматриваемого уравнения.

Реализацию этой программы мы начнем с исследования поведения u и v при заменах τ (§§ 10–13). Именно с этой частью программы связаны наиболее громоздкие вычисления, но их идейная сторона весьма проста.

§ 10. Преобразования порядка 5

Матрице $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ можно сопоставить дробно-линейное преобразование $\tau \mapsto \dfrac{a\tau + b}{c\tau + d}$. При этом матрице $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ соответствует преобразование $\tau \mapsto \tau + 1$, а матрице $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ — преобразование $\tau \mapsto -1/\tau$. Легко проверить, что матрице AB соответствует преобразование $\tau \mapsto A(B\tau)$, а матрице $-E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ соответствует тождественное преобразование. Формулы (8.1) и (8.2) можно записать следующим образом:

$$\eta(T\tau) = e^{\pi i/12}\eta(\tau), \tag{10.1}$$

$$\eta(S\tau) = \sqrt{-i\tau}\,\eta(\tau). \tag{10.2}$$

Пусть $SL_2(\mathbb{Z})$ — группа целочисленных матриц порядка 2, имеющих определитель 1. В § 18 мы докажем, что группа $SL_2(\mathbb{Z})/\pm E$ порождена элементами S и T. Это означает, что если $A \in SL_2(\mathbb{Z})$, то $\eta(A\tau)$ можно выразить через $\eta(\tau)$ с помощью формул (10.1) и (10.2). Такое явное выражение через элементы матрицы A впервые было получено Дедекиндом (именно поэтому функция $\eta(\tau)$ носит его имя). Аналогичное выражение можно получить и для $f(A\tau)$, но при этом может встретиться не только $f(\tau)$, но и $f_1(\tau)$, и $f_2(\tau)$. Например, $f(T\tau) = e^{-\pi i/24}f_1(\tau)$. Нам понадобятся такие выражения лишь для некоторых конкретных матриц A. Мы выведем их отдельно.

Решение уравнения 5-й степени основано на изучении преобразований $f(\tau) \mapsto f(P\tau)$, где $P = \begin{pmatrix} p & q \\ r & s \end{pmatrix}$ — целочисленная матрица с определителем 5. Если $A \in SL_2(\mathbb{Z})$, то $f(AP\tau)$ выражается через $f(P\tau)$, $f_1(P\tau)$ и $f_2(P\tau)$. Поэтому следует выяснить, к какому простейшему виду можно привести матрицу P умножением слева на матрицу $A \in SL_2(\mathbb{Z})$.

Пусть c и d — такие взаимно простые числа, что $cp + dr = 0$. Существуют такие целые числа a и b, что $ad - bc = 1$. Это означает, что

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \quad \text{и} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \begin{pmatrix} p' & q' \\ 0 & s' \end{pmatrix},$$

причем $p's' = 5$. Воспользовавшись тем, что

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}\begin{pmatrix} p & q \\ 0 & s \end{pmatrix} = \begin{pmatrix} p & q + ns \\ 0 & s \end{pmatrix},$$

матрицу P можно привести к виду $\begin{pmatrix} p & q \\ 0 & s \end{pmatrix}$ где $-s/2 \leqslant q \leqslant s/2$. В итоге получаем, что P можно привести к одному из следующих видов:

$$P_\infty = \begin{pmatrix} 5 & 0 \\ 0 & 1 \end{pmatrix},\ P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 5 \end{pmatrix},\ P_{\pm 1} = \begin{pmatrix} 1 & \pm 1 \\ 0 & 5 \end{pmatrix},\ P_{\pm 2} = \begin{pmatrix} 1 & \pm 2 \\ 0 & 5 \end{pmatrix}.$$

Следовательно, все функции $f(P\tau)$ и $f_i(P\tau)$, где $i = 1, 2$ и P — целочисленная матрица с определителем 5, могут быть выражены через

$$v_\infty(\tau) = f(P_\infty\tau), \quad v_c = f(P_c\tau),$$

$$f_i(P_\infty\tau), \quad f_i(P_c\tau), \quad c = 0, \pm 1, \pm 2, \quad i = 1, 2.$$

Попытаемся выяснить, как ведет себя $v_c, c = 0, \pm 1, \pm 2$ (соответственно v_∞) при замене τ на $B\tau$, $B \in SL_2(\mathbb{Z})$.

Функция $v_c(\tau) = v(P_c\tau)$ заменяется при этом на $v(P_cB\tau)$ (соответственно $v_\infty(\tau)$ — на $v(P_\infty B\tau)$). Подобрав матрицу $A \in SL_2(\mathbb{Z})$ так, чтобы матрица $Q = AP_cB$ (соответственно $AP_\infty B$) оказалась равна P_∞, P_0, $P_{\pm 1}$ или $P_{\pm 2}$, можно будет выразить $v_c(B\tau)$ (соответственно $v_\infty(B\tau)$) через $f(Q\tau)$ или $f_i(Q\tau)$. Следующие три параграфа посвящены вычислениям этих выражений для матриц B, соответствующих заменам τ на $\tau + 2$, $-1/\tau$ и $\dfrac{\tau - 1}{\tau + 1}$.

§ 11. Замена τ на $\tau + 2$

При замене τ на $\tau + 1$ функция $f(\tau)$ заменяется на $e^{-\pi i/24}f_1(\tau)$ (см. (8.3)). Чтобы не прибегать к функции f_1, ограничимся заменой τ на $\tau + 2$. При этом, например, $v_\infty = f(5\tau)$ переходит в

$$f(5\tau + 10) = e^{-5\pi i/12}f(5\tau) = e^{-5\pi i/12}v_\infty,$$

а $v_0 = f(\tau/5)$ — в

$$f\left(\frac{\tau + 2}{5}\right) = v_2(\tau).$$

Получились преобразования разного вида: в одном случае есть множитель $e^{-5\pi i/12}$, а в другом множитель отсутствует. Дело в том, что мы неудачно выбрали функции v_0, $v_{\pm1}$, $v_{\pm2}$ и v_∞. Можно выбрать их так, что все преобразования будут выглядеть одинаково.

Обозначим через P_c матрицу $\begin{pmatrix} 1 & c \\ 0 & 5 \end{pmatrix}$, $c \in \mathbb{Z}$. Ясно, что $v_c(\tau) = v(P_c\tau)$ для всех $c \in \mathbb{Z}$. Так же, как в § 10 показывается, что любую целочисленную матрицу с определителем 5 умножением слева на матрицу из $SL_2(\mathbb{Z})$ можно привести к одной из матриц

$$P_\infty,\quad P_{c_1},\quad P_{c_2},\quad P_{c_3},\quad P_{c_4},\quad P_{c_5},$$

где $c_1,\ldots,c_5$ — полная система вычетов по модулю 5. Это означает, что функция $f(P\tau)$ может быть выражена через

$$f(P_\infty\tau),\ f_i(P_\infty\tau),\ f(P_{c_j}\tau),\ f_i(P_{c_j}\tau),\quad i=1,2,\ j=1,\ldots,5.$$

Установим до конца этой главы обозначения

$$v_\infty = f(P_\infty\tau),\quad v_0 = f(P_0\tau),\quad v_1 = f(P_{96}\tau),$$
$$v_2 = f(P_{-48}\tau),\quad v_3 = f(P_{48}\tau),\quad v_4 = f(P_{-96}\tau).$$

Этот, на первый взгляд странный, выбор функций $v(\tau)$ объясняется следующим образом. Во-первых, числа 0, ±48, ±96 образуют полную систему вычетов по модулю 5. Во-вторых, функция f имеет период 48, следовательно существует всего 240 различных функций $f\left(\dfrac{\tau+c}{5}\right)$. Оказывается, множитель, о котором говорилось в начале этого параграфа, возникающий при преобразовании функции $f\left(\dfrac{\tau+c}{5}\right)$, зависит от остатка при делении c на 48.

В дальнейшем мы будем считать, что индекс c у функции v_c является вычетом по модулю 5. Например, под функцией $v_5(\tau)$ следует понимать не $f\left(\dfrac{\tau+5}{5}\right)$, а $v_0(\tau) = f\left(\dfrac{\tau}{5}\right)$. Такое соглашение позволяет очень просто описать поведение функции v при замене τ на $\tau+2$. Функция v_c заменяется на $e^{-5\pi i/12}v_{c+2}$, а v_∞ заменяется на $e^{-5\pi i/12}v_\infty$. В самом деле,

$$v_\infty(\tau+2) = f(5(\tau+2)) = e^{-5\pi i/12}f(5\tau) = e^{-5\pi i/12}v_\infty(\tau),$$

$$v_0(\tau+2) = f\left(\frac{\tau+2}{5}\right) = f\left(\frac{\tau+50-48}{5}\right) = e^{-5\pi i/12}f\left(\frac{\tau-48}{5}\right) = e^{-5\pi i/12}v_2(\tau).$$

Аналогично проверяются остальные случаи.

Удивительным образом преобразование индекса c совпадает с преобразованием параметра τ. В следующих двух параграфах мы убедимся, что это верно и для замен τ на $-1/\tau$ и τ на $\dfrac{\tau-1}{\tau+1}$.

§ 12. Замена τ на $-1/\tau$

При замене τ на $-1/\tau$ функция $v_\infty = f(5\tau)$ заменяется на

$$f\left(-\frac{5}{\tau}\right) = f\left(\frac{\tau}{5}\right) = v_0.$$

Легко проверить, что v_0 заменяется на v_∞. Таким образом, v_c заменяется на $v_{-1/c}$ при $c = 0,\infty$. Для остальных значений индекса c это тоже верно, но доказательство требует громоздких вычислений.

Чтобы выяснить как преобразуются функции v_c, где $c \neq \neq 0,\infty$, нам придется представлять матрицы из $SL_2(\mathbb{Z})$ в виде произведений матриц S и T. Делать это проще всего для матриц с небольшими элементами, поэтому мы воспользуемся тем, что

$$v_{\pm2} = f\left(\frac{\tau\mp48}{5}\right) = f\left(\frac{\tau\pm2}{5}\mp10\right) = e^{\mp5\pi i/12}f\left(\frac{\tau\pm2}{5}\right),$$

$$v_{\pm1} = f\left(\frac{\tau\pm96}{5}\right) = f\left(\frac{\tau\mp4}{5}\pm20\right) = e^{\mp5\pi i/6}f\left(\frac{\tau\mp4}{5}\right).$$

При заменах τ на $-1/\tau$ функции $v_{\pm2}$ и $v_{\pm1}$ заменяются на

$$e^{\mp5\pi i/12}f\left(\frac{\pm2\tau-1}{5\tau}\right)\quad\text{и}\quad e^{\mp5\pi i/6}f\left(\frac{\mp4\tau-1}{5\tau}\right).$$

Вычисления мы начнем с функции $f\left(\dfrac{\tau+48}{5}\right) = v_{-2}$. Для нее нужно привести к одному из шести основных видов дробно-

линейную функцию $\dfrac{-2\tau-1}{5\tau}$. Следуя описанному в § 10 алгоритму приведения, получаем соотношение

$$\begin{pmatrix} 2 & 1 \\ -5 & -2 \end{pmatrix}\begin{pmatrix} -2 & -1 \\ 5 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & 5 \end{pmatrix}.$$

Поэтому

$$f\left(\frac{-2\tau-1}{5\tau}\right) = f\left(A^{-1}\left(\frac{\tau-2}{5}\right)\right),$$

где

$$A^{-1} = \begin{pmatrix} 2 & 1 \\ -5 & -2 \end{pmatrix}^{-1} = -STS^2TS^{-2}S.$$

Следовательно,

$$f\left(\frac{-2\tau-1}{5\tau}\right) = f\left(ST^2ST^{-2}S\left(\frac{\tau-2}{5}\right)\right).$$

Легко проверить, что

$$f(ST^2ST^{-2}S\beta) = f(T^2ST^{-2}S\beta) = \\ = e^{-\pi i/12}f(ST^{-2}S\beta) = e^{-\pi i/12+\pi i/12}f(\beta) = f(\beta).$$

Таким образом, функция $f\left(\dfrac{\tau+48}{5}\right)$ заменяется на

$$e^{-5\pi i/12}f\left(\frac{\tau-2}{5}\right) = e^{-5\pi i/12}f\left(\frac{\tau+48}{5}-10\right) = f\left(\frac{\tau+48}{5}\right).$$

Аналогичные вычисления показывают, что функция $f\left(\dfrac{\tau-48}{5}\right)$ заменяется на $f\left(\dfrac{\tau-48}{5}\right)$. Следовательно, v_c заменяется на v_c при $c=\pm 2$. Ясно также, что $-1/c = c$ при $c=\pm 2$.

Для функций $f\left(\dfrac{\tau\pm 96}{5}\right)$ вычисления чуть более длинные. Так как

$$\begin{pmatrix} 4 & 3 \\ 5 & 4 \end{pmatrix}\begin{pmatrix} -4 & -1 \\ 5 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -4 \\ 0 & -5 \end{pmatrix}$$

и

$$\begin{pmatrix} 4 & 3 \\ 5 & 4 \end{pmatrix}^{-1} = STST^{-4}ST^{-1},$$

то

$$f\left(\frac{-4\tau-1}{5\tau}\right) = f\left(STST^{-4}ST^{-1}\left(\frac{\tau+4}{5}\right)\right).$$

Легко проверить, что

$$f(STST^{-4}ST^{-1}\beta) = f(TST^{-4}ST^{-1}\beta) = \\ = e^{-\pi i/24}f_1(ST^{-4}ST^{-1}\beta) = e^{-\pi i/24}f_2(T^{-4}ST^{-1}\beta) = \\ = e^{-9\pi i/24}f_2(ST^{-1}\beta) = e^{-9\pi i/24}f_1(T^{-1}\beta) = e^{-\pi i/3}f(\beta).$$

Поэтому функция $f\left(\dfrac{\tau+96}{5}\right)$ заменяется на

$$e^{-5\pi i/6-\pi i/3}f\left(\frac{\tau+4}{5}\right).$$

Ясно, что

$$f\left(\frac{\tau+4}{5}\right) = f\left(\frac{\tau-96}{5}+20\right) = e^{5\pi i/6}f\left(\frac{\tau-96}{5}\right).$$

А так как $\dfrac{5}{6}+\dfrac{1}{3}+\dfrac{5}{6} = 2$ и $e^{-2\pi i} = 1$, то в результате получаем, что $v_1 = f\left(\dfrac{\tau+96}{5}\right)$ заменяется на $f\left(\dfrac{\tau-96}{5}\right) = v_{-1}$. Аналогичные вычисления показывают, что v_{-1} заменяется на v_1.

Таким образом, при замене τ на $-1/\tau$ функция v_c заменяется на $v_{-1/c}$ при всех c.

§ 13. Замена τ на $\dfrac{\tau-1}{\tau+1}$

Докажем, что при замене τ на $\dfrac{\tau-1}{\tau+1}$ функция v_c заменяется на $-\sqrt{2}/v_d$, где $d = \dfrac{c-1}{c+1}$. Существенную роль при доказательстве этого играет соотношение (8.9):

$$f(\tau)f\left(\frac{\tau-1}{\tau+1}\right) = \sqrt{2}.$$

Для вычислений нам нужно получить соотношение вида:

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}x&y\\0&z\end{pmatrix}=\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}\begin{pmatrix}a&b\\0&d\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix}, \qquad (13.1)$$

где матрица $\begin{pmatrix}a&b\\0&d\end{pmatrix}$ задана, а матрица $\begin{pmatrix}\alpha&\beta\\\gamma&\delta\end{pmatrix}\in SL_2(\mathbb{Z})$ выбирается так, чтобы элемент (2,2) матрицы $\begin{pmatrix}x&y\\0&z\end{pmatrix}$ был действительно нулевым. После умножения обеих частей равенства (13.1) на матрицу $\begin{pmatrix}1&-1\\1&1\end{pmatrix}$ это условие приводит к соотношению

$$(-\alpha+\gamma)(a+b)+(\delta-\beta)d=0.$$

Наиболее просто получаются соотношения для v_0 и v_∞; они имеют вид

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}1&-4\\0&5\end{pmatrix}=\begin{pmatrix}1&-4\\0&1\end{pmatrix}\begin{pmatrix}5&0\\0&1\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix},$$

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}1&4\\0&5\end{pmatrix}=\begin{pmatrix}1&0\\-4&1\end{pmatrix}\begin{pmatrix}1&0\\0&5\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix}.$$

Следовательно,

$$\frac{\sqrt{2}}{f\left(\dfrac{\tau-4}{5}\right)}=e^{4\pi i/24}f\left(5\,\frac{\tau-1}{\tau+1}\right),$$

$$\frac{\sqrt{2}}{f\left(\dfrac{\tau+4}{5}\right)}=e^{-4\pi i/24}f\left(\frac{1}{5}\,\frac{\tau-1}{\tau+1}\right).$$

При выводе второй формулы мы воспользовались тем, что

$$f\left(\frac{\tau}{-4\tau+1}\right)=f\left(\frac{4\tau-1}{\tau}\right)=e^{-4\pi i/24}f(\tau).$$

Учитывая, что $e^{-\pi i}=-1$, получаем

$$f\left(5\,\frac{\tau-1}{\tau+1}\right)=\frac{-\sqrt{2}}{f\left(\dfrac{\tau+96}{5}\right)},\qquad f\left(\frac{1}{5}\,\frac{\tau-1}{\tau+1}\right)=\frac{-\sqrt{2}}{f\left(\dfrac{\tau-96}{5}\right)}.$$

Для v_1 и v_4 соотношения имеют вид

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}1&0\\0&5\end{pmatrix}=\begin{pmatrix}3&-1\\-2&1\end{pmatrix}\begin{pmatrix}1&1\\0&5\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix}, \qquad (13.2)$$

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}5&0\\0&1\end{pmatrix}=\begin{pmatrix}3&1\\2&1\end{pmatrix}\begin{pmatrix}1&-1\\0&5\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix}, \qquad (13.3)$$

Проведем вычисления для v_1. При замене τ на $\dfrac{\tau-1}{\tau+1}$ функция $v_1=f\left(\dfrac{\tau+96}{5}\right)$ заменяется на

$$f\left(\frac{\alpha+96}{5}\right)=f\left(\frac{\alpha+1}{5}+19\right)=$$

$$=e^{-19\pi i/24}f_1\left(\frac{\alpha+1}{5}\right)=e^{-19\pi i/24}f_1(\beta),$$

где $\alpha=\dfrac{\tau-1}{\tau+1}$ и $\beta=\dfrac{\alpha+1}{5}$. Соотношение (13.2) означает, что

$$\frac{\sqrt{2}}{f\left(\dfrac{\tau}{5}\right)}=f\left(\frac{3\beta-1}{-2\beta+1}\right).$$

Легко проверить, что

$$f\left(\frac{3\beta-1}{-2\beta+1}\right)=e^{\pi i/24}f_1\left(\frac{\beta}{-2\beta+1}\right)=$$

$$=e^{\pi i/24}f_2(2-1/\beta)=e^{5\pi i/24}f_1(\beta).$$

Поэтому

$$e^{-19\pi i/24}f_1(\beta)=-e^{-5\pi i/24}f_1(\beta)=-\frac{\sqrt{2}}{f\left(\dfrac{\tau}{5}\right)},$$

т. е. функция v_1 заменяется на $-\sqrt{2}/v_0$. Аналогичные вычисления показывают, что v_4 заменяется на $-\sqrt{2}/f(5\tau)=-\sqrt{2}/v_\infty$.

Для v_2 получаем соотношение

$$\begin{pmatrix}1&-1\\1&1\end{pmatrix}\begin{pmatrix}-1&-2\\0&-5\end{pmatrix}=\begin{pmatrix}-2&1\\3&-2\end{pmatrix}\begin{pmatrix}1&2\\0&5\end{pmatrix}\begin{pmatrix}1&-1\\1&1\end{pmatrix},$$

а затем проверяем, что v_2 переходит в $-\sqrt{2}/v_2$. Аналогично доказывается, что v_3 переходит в $-\sqrt{2}/v_3$.

Результаты вычислений, проведенных в §§ 11–13, можно записать в виде следующей таблицы, где $\varepsilon = e^{-5\pi i/12}$:

	u	v_∞	v_0	v_1	v_2	v_3	v_4
$\tau+2$	$e^{-\pi i/12}u$	εv_∞	εv_2	εv_3	εv_4	εv_0	εv_1
$-1/\tau$	u	v_0	v_∞	v_4	v_2	v_3	v_1
$\dfrac{\tau-1}{\tau+1}$	$\dfrac{\sqrt{2}}{u}$	$-\dfrac{\sqrt{2}}{v_1}$	$-\dfrac{\sqrt{2}}{v_4}$	$-\dfrac{\sqrt{2}}{v_0}$	$-\dfrac{\sqrt{2}}{v_2}$	$-\dfrac{\sqrt{2}}{v_3}$	$-\dfrac{\sqrt{2}}{v_\infty}$

§ 14. Функции, инвариантные относительно замен τ на $\tau+2$, $-1/\tau$ и $\dfrac{\tau-1}{\tau+1}$

Функция $f^{24}(\tau)$ сохраняется при заменах τ на $\tau+2$ и на $-1/\tau$, а при замене τ на $\dfrac{\tau-1}{\tau+1}$ она заменяется на $-\dfrac{2^{12}}{f^{24}(\tau)}$. Если же мы рассмотрим функцию

$$F(\tau) = f^{24}(\tau) + \frac{2^{12}}{f^{24}(\tau)} = q^{-1}\prod(1+q^{2k-1})^{24} + \\ + 2^{12}q\prod(1+q^{2k-1})^{-24} = q^{-1} + 24 + \ldots,$$

то она сохраняется и при замене τ на $\dfrac{\tau-1}{\tau+1}$. В самом деле, при этой замене $\dfrac{2^{12}}{f^{24}(\tau)}$ заменяется на $f^{24}(\tau)$.

Можно доказать, что если функция $g(\tau)$, мероморфная в верхней полуплоскости $\operatorname{Im}\tau > 0$, сохраняется при замене τ на

$\tau+2$, $-1/\tau$ и $\dfrac{\tau-1}{\tau+1}$, то при некотором ограничении $g(\tau) = R(F(\tau))$, где R — рациональная функция. Ограничение заключается в том, что функция $g(q)$, где $q = e^{\pi i\tau}$, должна быть мероморфной. Для мероморфной функции $g(\tau)$ это условие сводится к тому, что точка $q=0$ не является существенно особой, т. е. в разложении $g(q) = \sum\limits_{n=-\infty}^{\infty} c_n q^n$ для отрицательных n все коэффициенты c_n, кроме конечного их числа, нулевые. (Доказательству этого утверждения посвящены §§ 17–20). Кроме того, уравнение $F(\tau) = c$ разрешимо для всех комплексных $c \neq 0$ (это мы доказываем в § 20). Если $R(F(\tau))$ не обращается в бесконечность при всех τ, для которых $F(\tau) \neq \infty$, то R — многочлен. В самом деле, пусть R — дробь с непостоянным знаменателем. Тогда ее знаменатель обращается в нуль при некотором $F(\tau) \neq \infty$. Если же $R(F)$ не обращается в бесконечность и при $q=0$ (т. е. $F=\infty$), то R — константа.

Мы воспользуемся этими фактами сначала для вывода модулярного уравнения, а затем и для решения уравнения 5-й степени.

§ 15. Вывод модулярного уравнения

Рассмотрим функции

$$A_c = \left(\frac{u}{v_c}\right)^3 + \left(\frac{v_c}{u}\right)^3, \qquad A_\infty = \left(\frac{u}{v_\infty}\right)^3 + \left(\frac{v_\infty}{u}\right)^3,$$

$$B_c = (uv_c)^2 - \frac{4}{(uv_c)^2}, \qquad B_\infty = (uv_\infty)^2 - \frac{4}{(uv_\infty)^2}.$$

Из результатов §§ 11–13 следует, что они преобразуются следующим образом:

	A	B
$\tau+2$	$-A$	$-B$
$-1/\tau$	A	B
$\dfrac{\tau-1}{\tau+1}$	$-A$	$-B$

Отсюда следует, что функция

$$(A_\infty - B_\infty)(A_0 - B_0)(A_1 - B_1)\dots(A_4 - B_4)$$

не изменяется при всех указанных преобразованиях τ, поскольку при этих преобразованиях лишь переставляются сомножители.

В дальнейшем величины A_0 и B_0 будут обозначаться через A_5 и B_5 соответственно. С одной стороны, это не противоречит нашему соглашению, что индекс c понимается как вычет по модулю 5. С другой стороны, использование A_5 и B_5 традиционно (например, см. [Б11, б]) и нам бы не хотелось нарушать эту традицию.

Легко проверить, что

$$A_\infty = q^{-1/2}(1 - 2q + \dots), \qquad B_\infty = q^{-1/2}(1 - 2q + \dots).$$

Поэтому $A_\infty - B_\infty$ обращается в нуль при $q = 0$, т. е. при $\operatorname{Im}\tau \to \infty$. Покажем, что $A_c - B_c$ обращается в нуль при $q = 0$ для всех c. Ясно, что

$$u(5\tau - c) = f(5\tau) = v_\infty(\tau),$$

$$v_c(5\tau - c) = f\left(\frac{5\tau - c + c}{5}\right) = f(\tau) = u(\tau).$$

Кроме того, функции A и B не изменяются при замене u на v, а v на u. Поэтому

$$A_c(5\tau - c) = A_\infty(\tau), \qquad B_c(5\tau - c) = B_\infty(\tau). \tag{15.1}$$

А так как условия $\operatorname{Im}(5\tau - c) \to \infty$ и $\operatorname{Im}\tau \to \infty$ эквивалентны, то $A_c - B_c$ обращается в нуль при $q = 0$.

В результате мы получаем, что функция $\prod\limits_c (A_c - B_c)^2$ постоянна, причем при $q = 0$ она обращается в нуль. Следовательно, $A_c - B_c = 0$ для некоторого c. Формулы (15.1) показывают, что тогда $A_c - B_c = 0$ для всех c. Таким образом,

$$\left(\frac{u}{v}\right)^3 + \left(\frac{v}{u}\right)^3 = (uv)^2 - \frac{4}{(uv)^2},$$

т. е.

$$v^6 - u^5v^5 + 4uv + u^6 = 0. \tag{15.2}$$

Уравнение (15.2) связывает $u = f(\tau)$ и $v = f(5\tau)$; его называют *модулярным уравнением*. Если фиксировать $u = f(\tau)$ и рассматривать (15.2) как уравнение относительно v, то его корнями будут v_∞ и v_c, $c = 0, \pm1, \pm2$. Из теоремы Виета следует, в частности, что

$$\prod v_c = u^6. \tag{15.3}$$

§ 16. Решение уравнения 5-й степени

В предыдущем параграфе было показано, что коэффициенты многочлена 6-й степени с корнями v_c выражаются через u. Докажем теперь, что коэффициенты многочлена 5-й степени с корнями w_0, w_1, w_2, w_3 и w_4, где

$$w_z = \frac{(v_\infty - v_z)(v_{z+1} - v_{z-1})(v_{z+2} - v_{z-2})}{\sqrt{5}\,u^3}, \tag{16.1}$$

тоже выражаются через u, и найдем явный вид этого многочлена.

С помощью таблицы преобразований функции v_c можно получить следующую таблицу преобразований функций w_z:

	w_0	w_1	w_2	w_3	w_4
$\tau + 2$	$-w_2$	$-w_3$	$-w_4$	$-w_0$	$-w_1$
$-1/\tau$	w_0	w_2	w_1	w_4	w_3
$\dfrac{\tau - 1}{\tau + 1}$	$-w_0$	$-w_3$	$-w_4$	$-w_2$	$-w_1$

При этом для вычисления преобразований w_z при замене τ на $\dfrac{\tau - 1}{\tau + 1}$ нужно воспользоваться соотношением (15.3).

Рассмотрим многочлен

$$\prod(w - w_i) = w^5 + A_1w^4 + A_2w^3 + A_3w^2 + A_4w + A_5. \tag{16.2}$$

Его коэффициенты не обращаются в бесконечность при $u \neq 0, \infty$. Кроме того, функции A_1^2, A_2, A_3^2, A_4 и A_5^2 не изменяются при заменах τ на $\tau+2$, $-1/\tau$ и $\dfrac{\tau-1}{\tau+1}$. Поэтому они являются полиномами от $u^{24}+2^{12}u^{-24} = q^{-1}+24+\ldots$. Такой полином отличен от константы лишь в том случае, когда его разложение по степеням q начинается с члена, степень которого не превосходит -1.

Вычислим первый член разложения функции w_z. Так как

$$f(\tau) = q^{-1/24}\prod_{k=1}^{\infty}(1+q^{2k-1}),$$

где $q = e^{\pi i\tau}$, то первый член разложения v_c равен $(q')^{-1/24}$, где $q' = e^{5\pi i\tau}$ при $c=\infty$ и $q' = e^{\pi i(\tau+c)/5}$ при $c \neq \infty$. Легко проверить, что $e^{-\pi ic/120} = \alpha^c$, где $\alpha = e^{-4\pi i/5}$. Поэтому первый член разложения функции w_z равен

$$\frac{q^{-5/24}q^{-1/120}(\alpha^{z+1}-\alpha^{z-1})\,q^{-1/120}(\alpha^{z+2}-\alpha^{z-2})}{\sqrt{5}\,q^{-1/8}} = \lambda q^{-1/10},$$

где $\lambda = \alpha^{2z}(\alpha^3-\alpha-\alpha^{-1}+\alpha^{-3})/\sqrt{5} = \alpha^{2z}$, так как

$$\alpha^3-\alpha-\alpha^{-1}+\alpha^{-3} = 2(\cos 36^\circ+\cos 72^\circ) = \sqrt{5}.$$

Таким образом, разложение функции A_1 начинается с члена $q^{-1/10}$ (или с члена еще более высокой степени). Следовательно, функции A_1^2, A_2, A_3^2 и A_4 постоянны, а функция A_5^2 линейно зависит от $u^{24}+2^{12}u^{-24} = q^{-1}+24+\ldots$, поскольку ее разложение начинается с члена μq^{-1}, где $\mu = (\alpha^2\alpha^4\alpha^6\alpha^8)^2 = 1$. Сравнивая первые члены разложений функций $u^{24}+2^{12}u^{-24}$ и A_5^2, получаем

$$A_5^2 = u^{24}+\frac{2^{12}}{u^{24}}+C.$$

Чтобы вычислить константы C, A_1, A_2, A_2 и A_4, достаточно вычислить значения v_c для некоторого конкретного τ. Удобнее всего это сделать для $\tau = i$. В самом деле, $-1/i = i$, поэтому согласно (8.7) получаем

$$f_1(i) = f_2(i). \tag{16.3}$$

Ясно также, что для чисто мнимого τ функции f, f_1 и f_2 принимают положительные значения. Поэтому из (16.3), (6.1) и (6.2)

следует, что

$$u = f(i) = \sqrt[4]{2}$$

Воспользовавшись тем, что $(2-i)(2+i) = 5$, получаем

$$v_3 = f\left(\frac{i+48}{5}\right) = f\left(\frac{i-2}{5}+10\right) =$$

$$= e^{-10\pi i/24}f\left(\frac{i-2}{5}\right) = e^{-10\pi i/24}f(i+2) = e^{-\pi i/2}f(i) = -i\sqrt[4]{2}.$$

Аналогично $v_2 = i\sqrt[4]{2}$.

При $\tau = i$ модулярное уравнение принимает вид

$$v^6-a^5v^5+a^9v+a^6 = 0, \tag{16.4}$$

где $a = \sqrt[4]{2}$. Мы уже нашли два корня этого уравнения, а именно, $v_3 = -ia$ и $v_2 = ia$. Поделив многочлен (16.4) на $(v-v_2)(v-v_3) = v^2+a^2$, получим многочлен

$$v^4-a^5v^3+a^2v^2+a^7v+a^4 = (v-\alpha)^2(v-\beta)^2,$$

где $\alpha+\beta = a$ и $\alpha\beta = -a^2$. Если предположить, что $\alpha > 0$ и $\beta < 0$, то получим $\alpha = \dfrac{a(1+\sqrt{5})}{2}$ и $\beta = \dfrac{a(1-\sqrt{5})}{2}$. Ясно также, что

$$v_\infty = f(5i) = f\left(-\frac{1}{5i}\right) = f\left(\frac{i}{5}\right) = v_0,$$

причем $v_\infty > 0$, так как $f(\tau) > 0$ для любого чисто мнимого τ. Поэтому

$$v_0 = v_\infty = \frac{\sqrt[4]{2}(1+\sqrt{5})}{2}, \quad v_1 = v_4 = \frac{\sqrt[4]{2}(1-\sqrt{5})}{2}.$$

Подставив в (16.1) значения v_c при $\tau = i$, получим

$$w_0 = 0, \quad w_1 = w_2 = i\sqrt{5}, \quad w_3 = w_3 = -i\sqrt{5}.$$

Следовательно, при $\tau = i$ многочлен (16.2) имеет вид

$$w(w-i\sqrt{5})^2(w+i\sqrt{5})^2 = w(w^2+5)^2.$$

Таким образом, $A_5^2(i) = 0$, т. е.

$$C = -\left(u^{24}(i) + \frac{2^{12}}{u^{24}(i)}\right) = -(2^6 + 2^6) = -2^7.$$

Поэтому

$$A_5^2 = u^{24} + \frac{2^{12}}{u^{24}} - 2^7 = \left(u^{12} - \frac{2^6}{u^{12}}\right)^2,$$

т. е.

$$A_5^2 = \pm\left(u^{12} - \frac{2^6}{u^{12}}\right).$$

Легко проверить, что разложения функций $A_5 = -w_0w_1w_2w_3w_4$ и $u^{12} - 2^6u^{-12}$ начинаются с $-q^{1/2}$ и $q^{1/2}$ соответственно. Поэтому $A_5 = -u^{12} + 2^6u^{-12}$, а значит, уравнение (16.2) имеет вид

$$w(w^2+5)^2 = u^{12} - 64u^{-12}.$$

Из соотношений $f^8 = f_1^8 + f_2^8$ и $ff_1f_2 = \sqrt{2}$ следует, что

$$u^{12} - \frac{64}{u^{12}} = \frac{f^{24}(\tau) - 64}{f^{12}(\tau)} = \left(\frac{f_1^8(\tau) - f_2^8(\tau)}{f^2(\tau)}\right)^2.$$

Поэтому

$$\sqrt{w(\tau)} = \pm\frac{f_1^8(\tau) - f_2^8(\tau)}{f^2(\tau)\,(w^2(\tau)+5)}.$$

Положим

$$y(\tau) = \frac{f_1^8(\tau) - f_2^8(\tau)}{f^2(\tau)\,(w^2(\tau)+5)}. \tag{16.5}$$

Тогда

$$y^5 + 5y = y(w^2+5) = \frac{f_1^8 - f_2^8}{f^2}.$$

Если $\dfrac{f_1^8 - f_2^8}{f^2} = a$, то корни уравнения

$$y^2 + 5y = a \tag{16.6}$$

можно вычислить следующим образом. Сначала вычислим $v_c(\tau)$, $c = \infty, 0, 1, 2, 3, 4$. Затем по формуле (16.1) вычислим $w_z(\tau)$, $z = 0, 1, 2, 3, 4$. Наконец, по формуле (16.5) вычислим $y_z(\tau)$, $z = 0, 1, 2, 3, 4$. Это и есть корни уравнения (16.6).

Чтобы научиться решать уравнение (16.6) с любым параметром a, нужно лишь научиться решать уравнение

$$f_1^8(\tau) - f_2^8(\tau) = af^2(\tau) \tag{16.7}$$

Если учесть, что $f^8 = f_1^8 + f_2^8$ и $ff_1f_2 = \sqrt{2}$, то после возведения в квадрат это уравнение можно привести к виду

$$f^{24} - a^2f^{12} - 64 = 0. \tag{16.8}$$

Уравнение (16.8) является квадратным относительно f^{12}. Один его корень дает решение уравнения (16.7), а другой — решение уравнения, полученного из (16.7) заменой a на $-a$.

В § 14 мы сформулировали некоторые свойства функций, инвариантных относительно замен τ на $\tau + 2$, $-1/\tau$ и $\dfrac{\tau - 1}{\tau + 1}$. Этими свойствами мы воспользовались для решения уравнения 5-й степени. Займемся теперь их доказательством.

§ 17. Основная модулярная функция $j(\tau)$

В этом параграфе мы построим функцию $j(\tau)$, инвариантную относительно замен τ на $\tau + 1$ и $-1/\tau$. Функция $j(\tau)$ интересна тем, что любую функцию, инвариантную относительно этих замен и удовлетворяющую некоторым условиям мероморфности, можно представить в виде рациональной функции от j.

Рассмотрим сначала функции

$$k(\tau) = \frac{\theta_2^2(0\,|\,\tau)}{\theta_3^2(0\,|\,\tau)}, \quad k'(\tau) = \frac{\theta_0^2(0\,|\,\tau)}{\theta_3^2(0\,|\,\tau)}.$$

Они связаны соотношением $k^2 + k'^2 = 1$ (см. § 3). Полученные в § 7 формулы преобразований тэта-функций по параметру τ позволяют проверить, что

$$k(\tau+1) = i\frac{k(\tau)}{k'(\tau)}, \qquad k'(\tau+1) = \frac{1}{k'(\tau)},$$
$$k(-1/\tau) = k'(\tau), \qquad k'(-1/\tau) = k(\tau).$$

Рассмотрим теперь функцию $\lambda(\tau) = (k'(\tau))^2$. При заменах τ на $\tau+1$ и $-1/\tau$ величина λ переходит в $1/\lambda$ и $1-\lambda$ соответственно. Мы хотим получить функцию $j(\tau)$, инвариантную относительно этих замен. Для этого достаточно построить рациональную функцию от λ, инвариантную относительно замен λ на $1/\lambda$ и $1-\lambda$. Такую функцию можно получить следующим образом. Под действием указанных преобразований число λ может переходить лишь в следующие шесть чисел:

$$\lambda,\quad \frac{1}{\lambda},\quad 1-\lambda,\quad \frac{1}{1-\lambda},\quad \frac{\lambda}{1-\lambda},\quad \frac{\lambda-1}{\lambda}.$$

Поэтому функция

$$J_1(\lambda) = \lambda^2 + \frac{1}{\lambda^2} + (1-\lambda)^2 + \left(\frac{1}{1-\lambda}\right)^2 + \left(\frac{\lambda}{1-\lambda}\right)^2 + \left(\frac{\lambda-1}{\lambda}\right)^2$$

инвариантна относительно указанных преобразований. Часто бывают более удобны две другие функции, полученные из J_1 линейными преобразованиями, а именно,

$$J_2 = \frac{J_1+3}{2} = \frac{(\lambda^2-\lambda+1)^3}{\lambda^2(1-\lambda)^2},$$

$$J_3 = J_2 - \frac{27}{4} = \left(\frac{(\lambda+1)(\lambda-2)(\lambda-1/2)}{\lambda(1-\lambda)}\right)^2.$$

Для наших целей наиболее удобна функция J_2. Положим

$$j(\tau) = 2^8 J_2 = 2^8\frac{(\lambda^2-\lambda+1)^3}{\lambda^2(1-\lambda)^2} = 2^8\frac{(1-k^2k'^2)^3}{k^4k'^4} =$$

$$= 2^8\frac{(\theta_3^8-\theta_2^4\theta_0^4)^3}{\theta_0^8\theta_2^8\theta_3^8} = (2\pi)^8\frac{(\theta_3^8-\theta_2^4\theta_0^4)^3}{\theta_1'^8} =$$

$$= (f^{16}-f_1^8f_2^8)^3 = \frac{(f^{24}-16)^3}{f^{24}}.$$

Воспользовавшись тем, что $f^{24}(\tau) = q^{-1}\prod(1+q^{2k-1})^{24}$, для функции $j(\tau)$ можно получить следующее разложение:

$$j(\tau) = \left(f^{16}-\frac{16}{f^8}\right)^3 = q^{-2}\prod(1+q^{2k-1})^{48} - 48q^{-1}\prod(1+q^{2k-1})^{24} +$$

$$+3\cdot 16^2 - 16^3 q\prod(1+q^{2k-1})^{-24} = q^{-2} + 744 + \sum_{n=1}^{\infty} c_n q^n.$$

Функция $j(\tau)$ не изменяется при заменах τ на $\tau+1$ и $-1/\tau$. А так как при замене $\tau \mapsto \tau+1$ величина $q = e^{\pi i\tau}$ заменяется на $-q$, то функция $j(q)$ четная. Поэтому все коэффициенты c_n с нечетными номерами нулевые.

§ 18. Фундаментальная область функции $j(\tau)$

Функция $j(\tau)$ определена в верхней полуплоскости

$$H = \{\tau \in \mathbb{C} \mid \operatorname{Im}\tau > 0\}$$

В точках, получающихся друг из друга преобразованиями $\tau \mapsto \tau \pm 1$ и $\tau \mapsto -1/\tau$, значения функции j совпадают.

Покажем, что множество

$$D = \{\tau \in H \mid |\tau| \geqslant 1, |\operatorname{Re}\tau| \leqslant 1/2\}$$

обладает следующими двумя свойствами:

1) любую точку τ, $\tau \in H$, с помощью композиции преобразований $\tau \mapsto \tau \pm 1$ и $\tau \mapsto -1/\tau$ можно перевести в точку $\tau' \in D$;

2) никакие две различные внутренние точки множества D нельзя перевести друг в друга указанными преобразованиями.

Рассмотрим группу $G = SL_2(\mathbb{Z})/\pm E$ и группу G', порожденную преобразованиями S и T. Любой элемент группы G' имеет вид

$$\tau \mapsto \frac{a\tau+b}{c\tau+d},\quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

Поэтому можно считать, что G' — подгруппа G. Это замечание позволяет доказать, что для фиксированной точки $\tau \in H$ среди всех точек $g'\tau$, где $g' \in G'$, можно выбрать ту, у которой мнимая часть максимальна (и нельзя выбрать ту, у которой мнимая часть минимальна, поскольку $\lim\limits_{n\to\infty} \operatorname{Im}\left(\frac{-1}{\tau+n}\right) = 0$). Если $g \in G$, то

$$\operatorname{Im}(g\tau) = \operatorname{Im}\left(\frac{a\tau+b}{c\tau+d}\right) = \operatorname{Im}\frac{ad\tau+bc\bar{\tau}}{|c\tau+d|^2} = \frac{\operatorname{Im}\tau}{|c\tau+d|^2}.$$

Поэтому неравенство $\operatorname{Im}(g\tau) \geqslant \operatorname{Im}\tau$ эквивалентно неравенству $|c\tau+d| \leqslant 1$, которое выполняется лишь для конечного множества пар целых чисел (c,d). Таким образом, величина $\operatorname{Im}(g\tau)$ принимает конечное множество значений, не меньших $\operatorname{Im}\tau$ (хотя

каждое такое значение принимается для бесконечного множества элементов $g \in G$). В итоге получаем, что при поиске элемента $g \in G$, для которого мнимая часть числа $g\tau$ максимальна, можно ограничиться конечным набором элементов. Эти рассуждения справедливы не только для всей группы G, но и для любой ее подгруппы.

Пусть τ' — тот из образов точки $\tau \in H$ под действием группы G', у которого мнимая часть максимальна. Так как преобразования $\tau' \mapsto \tau' \pm 1$ не изменяют мнимую часть τ', то можно считать, что $|\operatorname{Re}\tau'| \leqslant 1/2$. Покажем, что в этом случае $\tau' \in D$, т. е. $|\tau'| \leqslant 1$. В самом деле, из условия $\operatorname{Im}\tau' \geqslant \operatorname{Im}(g'\tau')$ следует, в частности, что

$$\operatorname{Im}\tau' \geqslant \operatorname{Im}\left(-\frac{1}{\tau'}\right) = \frac{1}{|\tau'|^2}\operatorname{Im}\tau',$$

т. е. $|\tau'| \geqslant 1$.

Докажем теперь свойство 2, причем в уточненном виде. А именно, покажем, что под действием элементов группы G' граничные точки области D отождествляются так, как показано на рис. 41, т. е. под действием преобразований $\tau \mapsto \tau \pm 1$ склеиваются лучи QP и SP, дуги QR и SR тоже склеиваются под действием преобразования $\tau \mapsto -1/\tau$. Никаких других склеек граничных точек не происходит. Отметим, что если бесконечно удаленная точка P включена в область D, то в результате указанной склейки из D получается сфера, т. е. j можно рассматривать как функцию на сфере. В следующем параграфе будет показано, что j задает взаимно однозначное отображение этой сферы на $\mathbb{C} \cup \infty$.

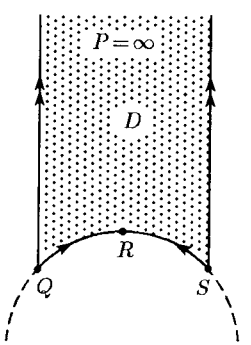


Рис. 41

Предположим, что $\tau' = g'\tau$, где $g' \in G'$, причем τ и τ' — две различные точки области D.

Можно считать, что

$$g'\tau = \frac{a\tau + b}{c\tau + d}, \quad \text{где} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}).$$

Рассмотрим сначала случай $c = 0$. В этом случае $ad = 1$, т. е. преобразование имеет вид $\tau \mapsto \tau \pm b$. Случай $b = 0$ приводит к совпадающим точкам τ и τ'. Если $b \neq 0$, то лишь для преобразований $\tau \mapsto \tau \pm 1$ образ множества D пересекается с самим множеством D, причем их пересечение принадлежит множеству $|\operatorname{Re}\tau| = 1/2$.

Предположим теперь, что $c \neq 0$. Тогда

$$\tau' = \frac{a\tau + b}{c\tau + d} = \frac{a}{c} - \frac{1}{c(c\tau + d)},$$

т. е.

$$|\tau' - a/c|\,|\tau + d/c| = 1/c^2. \tag{18.1}$$

Числа a/c и d/c вещественны, поэтому мнимые части чисел $\tau' - a/c$ и $\tau + d/c$ соответственно равны мнимым частям чисел τ' и τ, а значит, их модули не меньше $\sqrt{3}/2$. Следовательно, $|c| \leqslant 2/\sqrt{3}$. А так как c — ненулевое целое число, то $c = \pm 1$. Поэтому (18.1) можно записать в виде

$$|\tau' \mp a|\,|\tau \pm d| = 1.$$

Если $\tau \in D$ и $m \in \mathbb{Z}$, то $|\tau + m| \geqslant 1$, причем равенство возможно лишь в следующих случаях:

1) $|\tau| = 1$, $m = 0$;

2) $\tau = e^{\pi i/3}$, $m = 0, -1$;

3) $\tau = e^{2\pi i/3}$, $m = 0, 1$.

Таким образом, в случае $c \neq 0$ возможны склейки только тех точек, модуль которых равен 1, причем для точек, отличных от $e^{k\pi i/3}$, возможны склейки лишь по отображению $\tau \mapsto -1/\tau$. Для точек $e^{k\pi i/3}$ возможны склейки по отображениям $\tau \mapsto -1/\tau \pm 1$.

Попутно мы доказали еще одно уточнение свойства 2, а именно, если τ — внутренняя точка области D и $g\tau \in D$, где $g \in G'$, то g — тождественное преобразование. С помощью этого свойства можно доказать следующее утверждение.

Теорема 1. *Группа $G = SL_2(\mathbb{Z})/\pm E$ порождена элементами S и T, т. е. $G' = G$.*

Доказательство. Пусть g — произвольный элемент группы G. Рассмотрим некоторую внутреннюю точку τ_0 области D. Так как

$$\operatorname{Im}(g\tau) = \frac{\operatorname{Im}\tau}{|c\tau + d|^2},$$

то $g\tau_0 \in H$. Поэтому существует такой элемент g' группы G', что $g'(g\tau_0) \in D$. В результате мы получаем, что внутренняя точка τ_0

области D переводится в некоторую точку области D преобразованием $g'g \in G'$. Поэтому $g'g$ — тождественное преобразование, а значит, $g = (g')^{-1} \in G'$. □

Если мы хотим, чтобы в области D не было различных точек, получающихся друг из друга преобразованиями группы G, то этого можно добиться, например, исключив из области D луч PQ и дугу QR и оставив точки P, Q и R. Полученное множество называют *фундаментальной областью* группы G. Если не различать точки, получающиеся друг из друга преобразованиями S и T, то фундаментальной областью можно называть и саму область D. Фундаментальную область можно снабдить структурой топологического пространства, исходя из того, что она является фактор-пространством H по действию группы G.

Значение функции $j(\tau)$ в любой точке $\tau \in H$ совпадает со значением $j(\tau)$ в некоторой точке $\tau' \in D$. Кроме того, как будет показано в следующем параграфе, значения функции $j(\tau)$ в различных точках фундаментальной области группы G не могут совпадать. Поэтому D называют также фундаментальной областью функции $j(\tau)$.

§ 19. Решение уравнения $j(\tau) = c$

Функция $j(\tau) = q^{-2} + 744 + \ldots$ не обращается в бесконечность при $\operatorname{Im}\tau > 0$. Поэтому функция j не имеет полюсов в верхней полуплоскости H. При $\operatorname{Im}\tau = \infty$ функция $j(\tau)$ принимает бесконечное значение.

Решение уравнения $j(\tau) = c$ мы начнем с вычисления значений j в «вершинах» фундаментальной области, т. е. в точках i, $\varepsilon = e^{\pi i/3}$ и ε^2. Напомним, что

$$j(\tau) = \left(f^{16} - \frac{16}{f^8}\right)^3$$

и $f(i) = \sqrt[4]{2}$ (см. § 13). Поэтому $j(i) = 12^3 = 1728$. Для вычисления $j(\varepsilon) = j(\varepsilon)^2$ можно воспользоваться тем, что $\varepsilon = 1 - 1/\varepsilon$. В самом деле, из этого соотношения следует, что

$$f(\varepsilon) = f(1 - 1/\varepsilon) = e^{-\pi i/24} f_2(\varepsilon),$$
$$f_1(\varepsilon) = f_1(1 - 1/\varepsilon) = e^{-\pi i/24} f(\varepsilon).$$

Поэтому $\sqrt{2} = f(\varepsilon) f_1(\varepsilon) f_2(\varepsilon) = f^3(\varepsilon)$, а значит, $f^{24}(\varepsilon) = (\sqrt{2})^8 = 16$. Следовательно, $j(\varepsilon) = 0$.

Теорема 1. *Каждое значение c функция $j(\tau)$ принимает в фундаментальной области ровно один раз.*

Доказательство. Предположим сначала, что c отлично от 0 и 1728. Если на границе фундаментальной области принимается значение c, то фундаментальную область можно изменить так, как показано на рис. 42. Точнее говоря, если значение c принимается в некоторой точке границы, то мы вырезаем окрестность U этой точки и приклеиваем к D образ окрестности U при отображении S или $T^{\pm 1}$. Такую операцию нельзя проделать для «вершин», но они и не могут встретиться, поскольку $c \neq 0, 1728$.

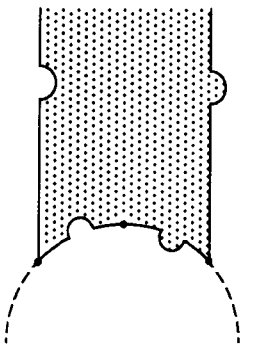


Рис. 42

Если мероморфная функция $g(z)$ не имеет полюсов в области G и не имеет нулей на ее границе ∂G, то количество нулей функции g, лежащих внутри области G, равно

$$\frac{1}{2\pi i} \int\limits_{\partial G} \frac{g'(z)}{g(z)}\, dz.$$

В самом деле, пусть $g(z) = c_0 (z-a)^r + c_1 (z-a)^{r+1} + \ldots$. Тогда

$$g'(z)/g(z) = r(z-a)^{-1} + \ldots.$$

Поэтому сумма вычетов особых точек функции g'/g, лежащих в области G, равна количеству нулей функции g с учетом их кратностей.

Возьмем в качестве G область, изображенную на рис. 43, а в качестве g возьмем функцию $j(z) - c$. Легко проверить, что

$$\int\limits_{\partial G} \frac{j'(z)\,dz}{j(z)-c} = \int\limits_{P_1}^{P_2} \frac{j'(z)\,dz}{j(z)-c} = \int\limits_{ia+1/2}^{ia-1/2} d\ln(j(z)-c).$$

В самом деле, $j(z+1)=j(z)$, поэтому интегралы по отрезкам P_1Q и SP_2 взаимно уничтожаются. Кроме того $j(-1/z)=j(z)$,

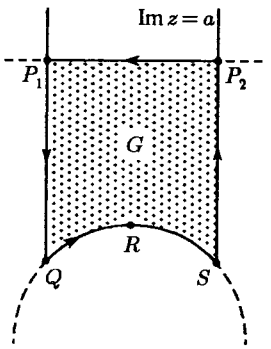


Рис. 43

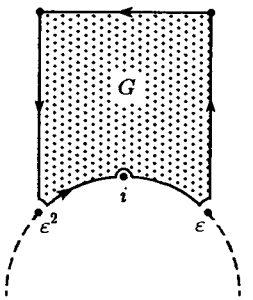


Рис. 44

поэтому интегралы по дугам QR и RS тоже взаимно уничтожаются. Это остается верным и в том случае, когда нам приходится изменять границу фундаментальной области.

Так как $j(\tau)=q^{-2}+744+\ldots=e^{-2\pi i\tau}+744+\ldots$, то при достаточно больших a для $z\in\left[ia-\frac{1}{2},ia+\frac{1}{2}\right]$ имеем

$$d\ln(j(z)-c)=d(-2\pi iz+\ldots)=-2\pi i\,dz+\ldots.$$

Следовательно,

$$\lim_{a\to\infty}\frac{1}{2\pi i}\int\limits_{ia+1/2}^{ia-1/2}d\ln(j(z)-c)=1.$$

В случае, когда $c=0$ или 1728, можно поступить следующим образом. Рассмотрим область G, изображенную на рис. 44. (Если на границе области принимается значение c, то область нужно подправить, как это делалось раньше.) Количество нулей функции $j(z)-c$, лежащих в области G, равно

$$\frac{1}{2\pi i}\int\limits_{\partial G}\frac{j'(z)\,dz}{j(z)-c}=1-\frac{r(i)}{2}-\frac{r(\varepsilon)}{3},$$

где $r(i)$ и $r(\varepsilon)$ — кратности нулей функции $j(z)-c$ в точках i и ε соответственно. Для доказательства этого утверждения достаточно заметить, что для точки i интеграл берется по половине

окружности, а для точек ε и ε^2 в сумме интеграл берется по трети окружности, причем окружность обходится по часовой стрелке, что приводит к знаку минус.

Если $c=0$ или 1728, то $r(\varepsilon)\neq 0$ или $r(i)\neq 0$. В этом случае количество нулей функции $j(\tau)-c$, лежащих в области G, строго меньше 1, т. е. $j(\tau)\neq c$ при $\tau\in G$. Отметим также, что если $c=0$, то $r(\varepsilon)=3$, а если $c=1728$, то $r(i)=2$. Это означает, что в точке ε функция $j(\tau)$ имеет нуль кратности 3, а в точке i функция $j(\tau)-1728$ имеет нуль кратности 2. □

§ 20. Функции, инвариантные относительно замен τ на $\tau+1$ и $-1/\tau$

Любая функция $g(\tau)$, определенная в верхней полуплоскости H и инвариантная относительно замен τ на $\tau+1$ и $-1/\tau$, представима в виде $G(j(\tau))$, где G — некоторая функция. В самом деле, как было показано в предыдущем параграфе, функция $j(\tau)$ задает взаимно однозначное отображение $D\cup\infty$ в $\mathbb{C}\cup\infty$. Поэтому существует обратное отображение j^{-1}: $\mathbb{C}\cup\infty\to D\cup\infty$. Положим $G(z)=g(j^{-1}(z))$. Тогда $G(j(\tau))=g(j^{-1}(j(\tau)))=g(\tau)$.

Предположим теперь, что функция g, рассматриваемая как функция от $q=e^{i\pi\tau}$, мероморфна в круге $|q|\leqslant 1$. Тогда G — рациональная функция. В самом деле, так как функция g мероморфна, то особые точки функции G в конечной области могут быть лишь полюсами. Значение $j=\infty$ соответствует $q=0$, поэтому в соответствии с условием точка ∞ не может быть существенно особой для функции G. Следовательно, функция G не имеет в $\mathbb{C}\cup\infty$ особых точек, отличных от полюсов. Вычтем из G главные части ее разложений в особых точках. В результате получим функцию, не имеющую особых точек в $\mathbb{C}\cup\infty$, т. е. константу. А так как все главные части разложений функции G в особых точках состоят из конечного числа членов, то G — рациональная функция.

§ 21. Функции, инвариантные относительно замен τ на $\tau+2$ и $-1/\tau$

Для решения уравнения 5-й степени нам потребовалось описание функций, инвариантных относительно замен τ на $\tau+2$,

$-1/\tau$ и $\dfrac{\tau-1}{\tau+1}$. Пока что мы описали лишь функции, инвариантные относительно замен τ на $\tau+1$ и $-1/\tau$. Но после этого уже легко получить описание интересующих нас функций.

Напомним, что $j(\tau)=(f^{24}(\tau)-16)^3/f^{24}(\tau)$. Функция $f^{24}(\tau)$ инвариантна относительно замен τ на $\tau+2$ и $-1/\tau$. В классе функций, инвариантных относительно этих замен, она играет ту же роль, какую играет функция $j(\tau)$ в классе функций, инвариантных относительно замен τ на $\tau+1$ и $-1/\tau$. А именно, любая функция g, инвариантная относительно замен τ на $\tau+2$ и $-1/\tau$, рационально выражается через f^{24} (в том случае, когда функция g, рассматриваемая как функция от $q=e^{i\pi\tau}$, мероморфна в круге $|q|<1$). Доказать это можно тем же способом, которым мы воспользовались в § 20. Для этого нужно доказать следующее утверждение.

Теорема 1. *Положим* $D_1=\{\tau\in H\,||\tau|\geqslant 1, |\operatorname{Re}\tau|\leqslant 1\}$. *Тогда*

а) *для группы* G_1, *порожденной преобразованиями* $\tau\mapsto\tau+2$ *и* $\tau\mapsto -1/\tau$, *множество* D_1 *является фундаментальной областью.*

б) *Каждое ненулевое значение функция* f^{24} *принимает в области* D_1 *ровно один раз.*

Доказательство. Пусть τ' — тот из образов точки $\tau\in H$ под действием группы G_1, у которого мнимая часть максимальна (существование такой точки τ' доказано в § 18). Так как преобразование $\tau'\mapsto\tau'\pm 2$ не изменяет мнимую часть τ', то можно считать, что $|\operatorname{Re}\tau'|\leqslant 1$. Покажем, что в этом случае $\tau'\in D_1$, т. е. $|\tau'|\geqslant 1$. В самом деле, по условию

$$\operatorname{Im}\tau'\geqslant\operatorname{Im}\left(\frac{1}{\tau'}\right)=\frac{1}{|\tau'|^2}\operatorname{Im}\tau',$$

поэтому $|\tau'|\geqslant 1$.

Мы показали, что любую точку $\tau\in H$ действием элемента $g\in G_1$ можно перевести в точку $\tau'\in D_1$. Чтобы доказать, что различные внутренние точки области D_1 нельзя перевести друг в друга действием элементов группы G_1, достаточно проверить утверждение б).

Пополним область D_1 (рис. 45) точками ± 1. Этим точкам соответствует $q=-1$. Для такого значения q получаем

$$f^{24}=q^{-1}\prod_{k=1}^{\infty}(1+q^{2k-1})^{24}=0,$$

причем этот нуль бесконечнократный, т. е. для функции $j(\tau)$ точки $\tau=\pm 1$ существенно особые. В области D_1 функция $j(\tau)$

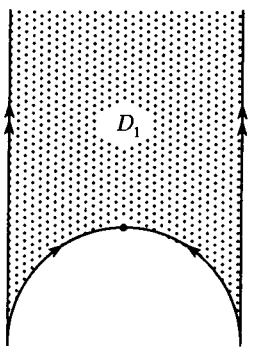


Рис. 45

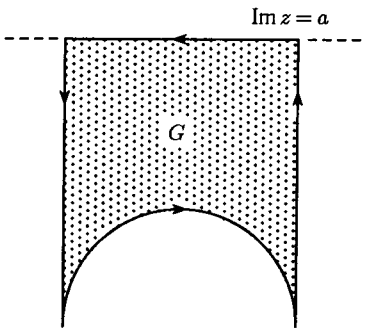


Рис. 46

не имеет полюсов, поэтому в области D_1 функция $f^{24}(\tau)$ не обращается в нуль. Кроме того, при $\operatorname{Im}\tau\to\infty$ функции $j(\tau)$ и $f^{24}(\tau)$ принимают значение ∞. Остается рассмотреть ненулевые значения функции $f^{24}(\tau)$.

Пусть G — область, изображенная на рис. 46. Тогда при $c\neq 0$, учитывая изображенные на рис. 45 склейки, получаем

$$\int\limits_{\partial G} d\ln(f^{24}(z)-c)=\int\limits_{ia+1/2}^{ia-1/2} d\ln(f^{24}(z)-c).$$

А так как $f^{24}(\tau)=q^{-1}+\ldots=e^{-\pi i\tau}+\ldots$, то

$$d\ln(f^{24}(z)-c)=d(-\pi iz+\ldots)=-\pi i\,dz+\ldots,$$

поэтому

$$\lim_{a\to\infty}\int\limits_{ia+1/2}^{ia-1/2} d\ln(f^{24}(z)-c)=1.$$

Это означает, что в области D_1 функция $f^{24}(\tau)$ принимает каждое ненулевое значение ровно один раз. □

Предположим теперь, что функция g, удовлетворяющая условию мероморфности, инвариантна не только относительно замен τ на $\tau+2$ и $-1/\tau$, но и относительно замены τ на $\dfrac{\tau-1}{\tau+1}$. Тогда $g(\tau)=R(f^{24}(\tau))$, где R — такая рациональная функция, что $R\left(f^{24}\left(\dfrac{\tau-1}{\tau+1}\right)\right)=R(f^{24}(\tau))$, т. е. $R\left(\dfrac{2^{12}}{f^{24}}\right)=R(f^{24})$.

Положим $f^{24}(\tau)=x$, $2^{12}=a$. Пусть

$$R(x)=\sum_{k=-\infty}^{\infty}c_kx^k.$$

Тогда

$$R\left(\frac{a}{x}\right)=\sum_{k=-\infty}^{\infty}c_ka^kx^{-k},$$

поэтому $c_ka^k=c_{-k}$, а значит,

$$R(x)=c_0+\sum_{k=1}^{\infty}c_k\left(x^k+\left(\frac{a}{x}\right)^k\right).$$

Рассматривая разности $\left(x+\dfrac{a}{x}\right)^k-\left(x^k+\left(\dfrac{a}{x}\right)^k\right)$, легко доказать индукцией по k, что $x^k+\left(\dfrac{a}{x}\right)^k$ полиномиально выражается через $x+\dfrac{a}{x}$. Таким образом, $R(x)=G\left(x+\dfrac{a}{x}\right)$, где G — некоторая функция. Так как функция R рациональна, то функция G не имеет особых точек, отличных от полюсов, поэтому и G — рациональная функция. В итоге получаем, что функция $g(\tau)$ рационально выражается через функцию $f^{24}(\tau)+\dfrac{2^{12}}{f^{24}(\tau)}$.

§ 22. Заключительные замечания

Приведенное нами решение уравнения 5-й степени следует в основном знаменитой книге Вебера [Б11, б]. У Вебера значительную часть изложения занимает мотивировка, поясняющая происхождение этого решения. А именно, он вычисляет группу

Галуа модулярного уравнения (это вычисление восходит к Галуа) и показывает, что уравнение 5-й степени в форме Бринга можно представить в виде резольвенты Галуа модулярного уравнения. Это означает, что решение уравнения 5-й степени сводится к решению модулярного уравнения, корни которого известны.

Как уже упоминалось, первое решение уравнения 5-й степени с помощью эллиптических модулярных функций опубликовал Эрмит в 1858 г. Тогда же аналогичные результаты получили Кронекер, Бриоски и Жубер. Об их исследованиях и о своем собственном подходе к решению уравнения 5-й степени в 1884 г. Клейн написал книгу «Лекции об икосаэдре и решении уравнения 5-й степени» [Б16]. Вебер писал свою книгу после них, поэтому его изложение просто и понятно.

Для корней общего уравнения n-й степени известно выражение через тэта-константы (но не от одной переменной, а от нескольких). Такое решение уравнения степени n содержится в статье Х. Умемуры, опубликованной как приложение к книге Д. Мамфорда «Лекции о тэта-функциях» [Б20].

Решение алгебраических уравнений можно получить не только с помощью тэта-функций, но и с помощью других специальных функций. Многочисленные исследования были посвящены решению алгебраических уравнений посредством гипергеометрических функций и их обобщений. Этот вопрос обсуждается в упомянутой книге Клейна. Пожалуй, наиболее глубокие результаты, относящиеся к этому направлению, содержатся в монографиях Лахтина [Б18]. Современный обзор можно найти в книге Белардинелли [Б4].

СПИСОК ЛИТЕРАТУРЫ

А. Классические работы

1. Абель Н. Х. (Abel N. H.)
Œuvres complétes du Niels Henrik Abel. — Christiania, 1881.

2. Бернулли И. (Bernoulli I.)
Opera Omnia. T. I–IV. — Lausannae; Genevae: Bosquet, 1742.

3. Гаусс К. Ф. (Gauss C. F.)
а) Труды по теории чисел. — М.: АН СССР, 1959.
б) Werke. Bd. 10. — Göttingen, 1917.

4. Диофант (Diophant)
Арифметика и книга о многоугольных числах. — М.: Наука, 1974.

5. Жордан К. (Jordan C.)
Traité des substitutions et des équations algébriques. — Paris, 1870.

6. Лагранж Ж. Л. (Lagrange J. L.)
Réflexions sur la résolution algébrique des équations // Mém. Acad. Berlin. — 1770–71.

7. Лежандр А. М. (Legendre A. M.)
Traité des fonctions elliptiques et des intégrales euleriennes. V. 1–3. — Paris, 1825–1832.

8. Пуанкаре А. (Poincaré H.)
Sur les propriétés des courbes algébriques planes // J. Liouville. — 1901. — V. 7. — P. 161–233.

9. Руффини П. (Ruffini P.)
Theoria generale delle equazioni in cui si dimostra impossibili la soluzione algebraica delle equazioni generali di grado superiore al quarta. — Bologna, 1799.

10. Серре Ж. А. (Serret J. A.)
а) Mémoire sur la représentation géométrique des fonctions elliptiques et ultra-elliptiques // J. math. pures et appl. — 1845. — V. 10. — P. 257.
б) Dévelopments sur une classe d'équations relatives à la représentation géométrique des fonctions elliptiques // J. math. pures et appl. — 1845. — V. 10. — P. 357.
в) Note sur les courbes elliptiques de la première espèce // J. math. pures et appl. — 1845. — V. 10. — P. 421.

11. Фаньяно Дж. (Fagnano G. C.)
Opere Matematiche del Marchese Guilio Carlo de'Toschi di Fagnano. V. 1–3. — Publ. dai Soci V. Volterra, G. Loria, D. Gamboli, 1911–1913.

12. Чирнгауз Э. (Tschirnhausen E.)
Nova metodus auferendi omnes terminos intermedios ex data aequatione // Acta eruditoru. — T. 2. — Leipzig, 1689.

13. Эйлер Л. (Euler L.)
а) Интегральное исчисление. Т. I. — М.: Гостехиздат, 1956.
б) Opera Omnia, Ser. I. T. XXI. — Leipzig; Berlin: Teubner, 1913.

14. Эрмит Ш. (Hermite Ch.)
а) Sur la résolution de l'équation du cinquième degré // C. r. Acad. Sci. Paris. — 1858. — V. 46.
б) Sur l'équation du cinquième degré // C. r. Acad. Sci. Paris. — 1865. — V. 61; 1866. — V. 62.

15. Якоби К. Г. (Jacobi C. G. J.)
Gesammelte Werke. — Berlin, 1881–1891.

Б. Основные руководства

1. Айерлэнд К., Роузен М. И. (Ireland K., Rosen M.)
Классическое введение в современную теорию чисел. — М.: Мир, 1987.

2. Алгебраическая теория чисел/Ред. Дж. Касселс, А. Фрёлих. — М.: Мир, 1969.

3. Ахиезер Н. И.
Элементы теории эллиптических функций. — М.: Наука, 1970.

4. Белардинелли Дж. (Belardinelli G.)
Fonctions hypergéométriques de plusieurs variables et résolution analitique des équations algébriques générales. — Paris: Gauthier-Villars, 1960.

5. Берже М. (Berger M.)
Геометрия. — М.: Мир, 1984.

6. Бернсайд У. С., Пантон А. У. (Burnside W. S., Panton A. W.)
Theory of equation. V. 1, 2. — Dublin, 1928.

7. Боревич З. И., Шафаревич И. Р.
Теория чисел. — М.: Наука, 1995.

8. Брискорн Э., Кнёррер Х. (Brieskorn E., Knörrer H.)
Plane algebraic curves. — Birkhäuser, 1986.

9. Бурбаки Н. (Bourbaki N.)
Алгебра. Многочлены и поля. Упорядоченные группы. — М.: Наука, 1965.

10. Ващенко-Захарченко М. Е.
Высшая алгебра. Теория подставлений и ее приложения к алгебраическим уравнениям. — Киев: Типогр. Имп. ун-та Св. Владимира, 1890.

11. Вебер Г. (Weber H.)
а) Lehrbuch der Algebra. Bd. 2. Gruppen. Lineare Gruppen. Anwendungen der Gruppen Theorie. Algebraische Zahlen. — Braunschweig, 1899.
б) Lehrbuch der Algebra. Bd. 3. Elliptische Funktionen und algebraische Zahlen. — Braunschweig, 1908.

12. Вейль А. (Weil A.)
Эллиптические функции по Кронекеру и Эйзенштейну. — М.: Мир, 1978.

13. Граве Д. А.
Элементы высшей алгебры. — Киев: Типогр. Имп. ун-та Св. Владимира, 1914.

14. Гриффитс Ф., Харрис Дж. (Griffiths Ph., Harris J.)
Принципы алгебраической геометрии. — М.: Мир, 1982.

15. Гурвиц А., Курант Р. (Hurwitz A., Courant R.)
Теория функций. — М.: Наука, 1968.

16. Клейн Ф. (Klein F.)
Лекции об икосаэдре и решении уравнений пятой степени. — М.: Наука, 1989.

17. Коблиц Н. (Koblitz N.)
Введение в эллиптические кривые и модулярные формы. — М.: Мир, 1988.

18. Лахтин Л. К.
а) Алгебраические уравнения, разрешимые в гипергеометрических функциях. — М., 1893.
б) Дифференциальные резольвенты алгебраических уравнений высших степеней. — М., 1896.

19. Ленг С. (Lang S.)
а) Эллиптические функции. — М.: Наука, 1984.
б) Основы диофантовой геометрии. — М.: Мир, 1986.

20. Мамфорд Д. (Mumford D.)
Лекции о тэта-функциях. — М.: Мир, 1988.

21. Морделл Л. Дж. (Mordell L. J.)
Diophantine equations. — New York: Academic Press, 1969.

22. Постников М. М.
Теория Галуа. — М.: Физматлит, 1963.

23. Роберт А. (Robert A.)
Elliptic curves // Lecture Notes in Math. — Springer, 1973. — V. 326.

24. Сальмон Дж. (Salmon G.)
A traite on higher plaine curves: intend as a sequel to a traition on conic. — Dublin: Smith, 1852.

25. Серре Ж. А. (Serret J. A.)
Cours de calcul différentiel et intégral. V. 1–3. — Paris: Gautier-Villars, 1879 (2^{me} ed.).

26. Силверман Дж. Х. (Silverman J. H.)
а) The arithmetic of elliptic curves. — Springer, 1986.
б) Advanced topics in the arithmetic of elliptic curves. — Springer, 1994.

27. Степанов С. А.
Арифметика алгебраических кривых. — М.: Наука, 1991.

28. Уиттекер Э. Т., Ватсон Дж. Н. (Whittaker E. T., Watson G. N.)
Курс современного анализа. Т. 1, 2. — М.: Физматгиз, 1963.

29. Уокер Р. (Walker R. J.)
Алгебраические кривые. — М.: ИЛ, 1952.

30. Фрике Р. (Fricke R.)
Die elliptischen Funktionen und ihre Anwendungen. Bd. 1, 2. — Leipzig; Berlin: Teubner, 1916–1922.

31. Хьюзмоллер Д. (Husemöller D.)
Elliptic curves. — Springer, 1987.

32. Чеботарев Н. Г.
Теория Галуа. — Гостехиздат, 1936.

33. Эрмит Ш.
Курс анализа. — Гостехиздат, 1936.

B. Избранные статьи

1. Бёрч Б. Дж. (Birch B. J.)
Conjectures on elliptic curves // Theory of Numbers/Proc. of Symp. in Pure Math. — Pasadena, 1963 — V. 8.

2. Бёрч Б. Дж., Суиннертон-Дайер Х. П. Ф. (Birch B. J., Swinnerton-Dyer H. P. F.)
Notes on elliptic curves // J. Reine und Angew. Math. — 1963. — Bd. 212. — S. 7–25; 1965. — Bd. 218. — S. 79–108.

3. Вейль А. (Weil A.)
Sur un théorème de Mordell // Bull. Sci. Math. 2. — 1930. — V. 54. — P. 182–191.

4. Грюневальд Ф. Й., Циммерт Р. (Grünewald F. J., Zimmert R.)
Über einige rationale elliptische Kurven mit freiem Rang $\geqslant 8$ // J. Reine und Angew. Math. — 1977. — Bd. 296. — S. 100–107.

5. Даммит Д. С. (Dummit D. S.)
Solving solvable quintics // Math. Comp. — 1991. — V. 57. — P. 387–401.

6. Игуса Дж. (Igusa J.)
On the transformation theory of elliptic functions // Amer. J. Math. — 1959. — V. 81. — P. 436–452.

7. Йел П. Б. (Yale P. B.)
Automorphisms of the complex numbers // Math. Mag. — 1966. — V. 39. — P. 135–141.

8. Касселс Дж. В. С. (Cassels J. W. S.)
Diophantine equations with special reference to elliptic curves // J. London Math. Soc. — 1966. — V. 41 — P. 193–291. [Имеется перевод: Диофантовы уравнения со специальным рассмотрением эллиптических кривых // Математика. Сб. переводов. 1968. — Т. 12:1. — С. 113–160; Т. 12:2. — С. 3–48.]

9. Коутс Дж., Уайлс Э. (Coates J., Wiles A.)
On the conjecture of Birch and Swinnerton-Dyer // Invent. Math. — 1977. — V. 39, № 3. — P. 223–251.

10. Кэли А. (Cayley A.)
On a new auxilary equation in the theory of equations of the first order // Phil. Trans. R. Soc. — London, 1861. — V. 91, P. 263–276.

11. Мазур Б. (Mazur B.)
Rational points on modular curves // Lecture Notes in Math. — Springer, 1976. — V. 601.

12. Мельников И. Г.
Задача деления лемнискаты // Уч. зап. ЛГПИ — 1958. — Т. 197. — С. 20–42.

13. Морделл Л. Дж. (Mordell L. J.)
а) On the rational solutions of the indeterminate equations of the third and fourth degrees // Proc. Cambr. Phil. Soc. — 1922. — V. 21. — P. 179–192.
б) The infinity of rational solutions of $y^2 = x^3 + k$ // J. London Math. Soc. — 1966. — V. 41. — P. 523–525.
в) On the magnitude of the integral solutions of the equation $ax^2 + by^2 + cz^2 = 0$ // J. Number Theory. — 1969. — V. 1. — P. 1–3.

14. Роузен М. И. (Rosen M. I.)
Abel's theorem on the lemniscate // Amer. Math. Monthly. — 1981. — V. 88. — P. 387–395.

15. Спирман Б. К., Вильямс К. С. (Spearman B. K., Williams K. S.)
Characterization of solvable quintics $x^5 + ax + b$ // Amer. Math. Monthly. — 1994. — V. 101. — P. 986–992.

16. Сундарарджан Т. (Soundararajan T.)
On the automorphisms of the complex number field // Math. Mag. — 1967. — V. 40. — P. 213.

17. Тейлор Р., Уайлс Э. (Taylor R., Wiles A.)
Ring-theoretic properties of certain Hecke algebras // Ann. Math. — 1995. — V. 141. — P. 552–613.

18. Тейт Дж. (Tate J.)
Arithmetic of elliptic curves // Invent. Math. — 1974. — V. 23. — P. 179–206.

19. Уайлс Э. (Wiles A.)
Modular elliptic curves and Fermat's Last theorem // Ann. Math. — 1995. — V. 141. — P. 443–551.

20. Хассе Г. (Hasse H.)
Abstrakte Begründung der komplexen Multiplikation und Riemannsche Vermutung in Funktionenkörpern // Abh. Math. Sem. Hamburg. — 1934. — Bd. 10. — S. 325–348.

21. Хольцер Л. (Holzer L.)
Minimal solutions of Diohpantine equations // Can. J. Math. — 1950. — V. 11. — P. 238–244.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ