По мотивам задачника

О преобразовании Гаусса — Ландена*

Г. Б. Шабат

§ 0. Введение

В 2017 г. в сборнике «Математическое просвещение» была поставлена задача 21.11 (см. [5, с. 273]), которую мы приводим в чуть-чуть изменённых обозначениях:

(★) *Пусть a, b — положительные вещественные числа. Их арифметическое и геометрическое средние определяются как*

$$a_1 := \frac{a+b}{2}, \quad b_1 := \sqrt{ab}.$$

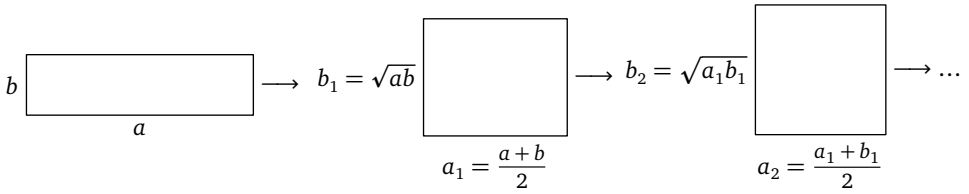
Докажите, что

$$\int_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+a^2)(u^2+b^2)}} = \int_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+a_1^2)(u^2+b_1^2)}}.$$

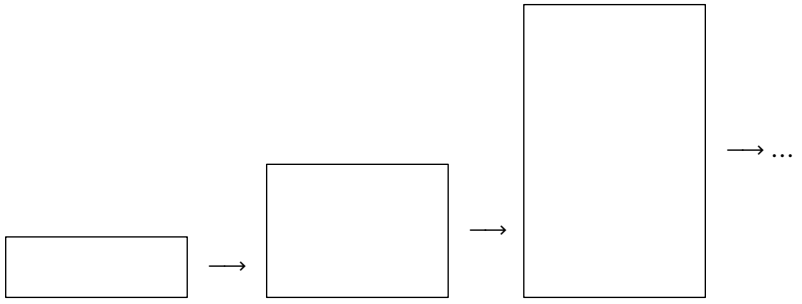
Предложенная для доказательства формула — классический результат, относящийся к математике конца XVIII – начала XIX века. Эта формула, её частные случаи и следствия из неё были предметом многолетних исследований Лагранжа и Гаусса — см. [2]. В настоящей статье мы обсудим формулу и связанную с ней математику, широко пользуясь средствами нашего времени.

* Джон Ланден (1719–1790) — английский математик. В 1755 году он выразил длину дуги гиперболы как сумму двух длин дуг эллипса, что впоследствии привело к *преобразованию Ландена*, связанному с материалом настоящей статьи.

Упомянем постановку геометрической задачи, связанной с одновременным рассмотрением арифметического и геометрического средних: *построить квадрат, наиболее близкий к заданному прямоугольнику*. Эту задачу можно уточнить двумя способами: *построить квадрат с тем же* периметром, *что заданный прямоугольник*, и *построить квадрат с той же* площадью, *что заданный прямоугольник*. Смешав эти два уточнения и *итерируя*, получим последовательность прямоугольников, очень быстро становящихся неотличимыми от квадратов[1)]:



В последующих параграфах будет приведено полное решение задачи (★), по существу заимствованное из дневников Гаусса. Однако *полное понимание* этого решения будет основано на связи приведённой последовательности прямоугольников с другой, в которой вертикальные стороны прямоугольников последовательно удваиваются[2)]:



Вышеуказанная последовательность итераций и объясняет интерес к рассматриваемой конструкции. Введя начальную пару положительных вещественных чисел $(a_0, b_0) := (a, b)$, рассмотрим две последовательности

$$a_{n+1} := \frac{a_n + b_n}{2}, \quad b_{n+1} := \sqrt{a_n b_n}.$$

[1)] Пример (в пикселях, $a = a_0$, $b = b_0$):

n	a_n	b_n
0	80	20
1	50	40
2	45	44,721…
3	44,860…	44,860…

[2)] Настоящая работа завершается изложением этой конструкции удвоения.

Нетрудно доказать (читатель может сделать это самостоятельно или обратиться к [2]), что эти последовательности имеют общий предел, который называется *арифметико-геометрическим средним* чисел a, b и обозначается[3)]

$$\mathrm{agM}(a,b) := \lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n.$$

Обе последовательности сходятся необычайно быстро. И во времена Лагранжа и Гаусса, и в наше время их пределы используются для приближённого вычисления интегралов, которым посвящена настоящая работа. (Подобные интегралы называются *эллиптическими*, поскольку возникли при вычислении длины дуги эллипса. Начиная с XVIII века они широко применяются и в математике, и в её разнообразных приложениях.) Дело в том, что, согласно (★), в очевидных обозначениях

$$\mathrm{agM}(a,b) = \mathrm{agM}(a_1,b_1) = \mathrm{agM}(a_2,b_2) = \ldots = \mathrm{agM}(a_\infty,b_\infty),$$

и

$$\int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+a_\infty^2)(u^2+b_\infty^2)}} = \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{u^2+a_\infty^2} = \frac{\pi}{a_\infty},$$

откуда получаем связь между арифметико-геометрическим средним и эллиптическим интегралом:

$$\mathrm{agM}(a,b) = \frac{\pi}{\displaystyle\int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+a^2)(u^2+b^2)}}}.$$

Поскольку левая часть этого равенства (по крайней мере с использованием современных компьютеров) легко вычисляется с любой точностью, знание достаточно хороших приближений числа π позволяет приближённо вычислять *произвольные* эллиптические интегралы. И наоборот, эта же формула позволяет на основе знания *некоторых* эллиптических интегралов с потрясающей точностью вычислять π — см. монографию [1].

§ 1. Численные эксперименты

Обсуждаемый нами факт — равенство двух положительных вещественных чисел; современные технологии позволяют быстро проверить его с любой точностью. Например, положив, как в рассмотренном выше

[3)] От слов *arithmetic-geometric Mean*.

примере, $a_0 = 2$, $b_0 = 18$ (чтобы следующая пара значений $a_1 = 6$, $b_1 = 10$ была рациональна), с помощью программы MAPLE получаем

$$\tfrac{1}{9}\mathrm{EllipticK}\Big(\tfrac{4}{9}\sqrt{5}\Big) = \tfrac{1}{5}\mathrm{EllipticK}\Big(\tfrac{4}{5}\Big).$$

Это равенство мало что скажет читателю, незнакомому с эллиптическими функциями и, в частности, с функцией EllipticK; однако сам факт выдачи такого результата говорит о том, что рассматриваемые интегралы присутствуют в библиотеке *специальных функций* этой программы.

Правда, если попросить MAPLE указать численные значения вычисленных интегралов, то на экране появится «равенство»

$$,3990605528 = ,3990605556,$$

т. е. совпадут только 8 десятичных знаков из 10.

Если, однако, настроить MAPLE на вычисления с 20 десятичными знаками, то «равенство» уточнится:

$$,39906055553294587753 = ,39906055553294587754.$$

Теперь совпадают 19 знаков из 20, и сомнения развеиваются.

Особую роль в истории математики сыграло вычисление Гаусса, датированное 30 мая 1799 года (см. [3, III, с. 361–371]). В наших обозначениях оно соответствует значениям $a = \sqrt{2}$, $b = 1$:

$$\int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+2)(u^2+1)}} = \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+1{,}207106\ldots^2)(u^2+1{,}189207\ldots^2)}} =$$
$$= \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+1{,}198156\ldots^2)(u^2+1{,}198123\ldots^2)}} =$$
$$= \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+1{,}198140\ldots^2)(u^2+1{,}198140\ldots^2)}} \approx$$
$$\approx \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{u^2+1{,}198140\ldots^2} = \frac{\pi}{1{,}198140\ldots} = 2{,}6220575542921198103\ldots$$

Во времена Гаусса это число было хорошо известно, и для него было введено специальное обозначение:

$$2{,}6220575542921198103\ldots =: \varpi.$$

Оно[4)] представляет собой *альтернативное* π. Дело в том, что со времён Эйлера эти два числа использовались параллельно:

$$\pi := \int\limits_{-1}^{1} \frac{\mathrm{d}x}{\sqrt{1-x^2}} = 3{,}141592\ldots, \quad \varpi := \int\limits_{-1}^{1} \frac{\mathrm{d}x}{\sqrt{1-x^4}} = 2{,}622057\ldots$$

Число 1,198140…, которое во введении получило обозначение $\mathrm{agM}(\sqrt{2}, 1)$, оказалось удовлетворяющим равенству

$$\mathrm{agM}(\sqrt{2}, 1) = \frac{\pi}{\varpi}.$$

В своём дневнике Гаусс пишет, связав обнаруженное равенство с проведёнными им ранее вычислениями длины лемнискаты: *доказательство этого факта несомненно откроет совершенно новую область анализа*.

Предсказание сбылось! Мы расскажем об этом в § 3 настоящей статьи.

§ 2. Подстановка, подсмотренная у Гаусса

В этом параграфе опять a и b — вещественные числа, удовлетворяющие неравенствам

$$a \geqslant b > 0, \tag{2.0}$$

и по ним строятся новые два:

$$a_1 = \frac{a+b}{2}, \quad b_1 = \sqrt{ab}. \tag{2.1}$$

Предложение. *Подстановка*

$$u = \sqrt{\frac{u_1^2 + b_1^2}{u_1^2 + a_1^2}}\, u_1 \tag{2.2}$$

устанавливает требуемое равенство:

$$\int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u}{\sqrt{(u^2+a^2)(u^2+b^2)}} = \int\limits_{-\infty}^{\infty} \frac{\mathrm{d}u_1}{\sqrt{(u_1^2+a_1^2)(u_1^2+b_1^2)}}. \tag{2.3}$$

Доказательство. Начнём с дифференциалов.

Лемма 1. *Выполнено равенство*

$$\mathrm{d}u = \frac{u_1^4 + 2a_1^2u_1^2 + a_1^2b_1^2}{u_1^2 + a_1^2} \frac{\mathrm{d}u_1}{\sqrt{(u_1^2+a_1^2)(u_1^2+b_1^2)}}. \tag{2.4}$$

[4)] В TEX-е эта буква ϖ называется *varpi* (варпи). Подобно этому, существуют $\vartheta, \varrho, \varphi$ наряду с θ, ρ, ϕ.

Доказательство. Действительно, согласно стандартным формулам

$$\mathrm{d}u =_{(2.2)} \mathrm{d}\left(\left(\frac{u_1^2+b_1^2}{u_1^2+a_1^2}\right)^{1/2} u_1\right) = u_1\,\mathrm{d}\left(\left(\frac{u_1^2+b_1^2}{u_1^2+a_1^2}\right)^{1/2}\right) + \left(\left(\frac{u_1^2+b_1^2}{u_1^2+a_1^2}\right)^{1/2}\right)\mathrm{d}u_1 =$$

$$= \frac{1}{2}u_1\left(\left(\frac{u_1^2+b_1^2}{u_1^2+a_1^2}\right)^{-1/2}\right)\mathrm{d}\left(\frac{u_1^2+b_1^2}{u_1^2+a_1^2}\right) + \sqrt{\frac{u_1^2+b_1^2}{u_1^2+a_1^2}}\,\mathrm{d}u_1.$$

Воспользовавшись формулой

$$\mathrm{d}\left(\frac{\alpha x+\beta}{\gamma x+\delta}\right) = \frac{\alpha\delta-\beta\gamma}{(\gamma x+\delta)^2}\,\mathrm{d}x$$

(здесь подразумевается, что $\alpha, \beta, \gamma, \delta$ постоянны), получим

$$\mathrm{d}u = \frac{1}{2}u_1\left(\left(\frac{u_1^2+a_1^2}{u_1^2+b_1^2}\right)^{1/2}\right)\frac{a_1^2-b_1^2}{(u_1^2+a_1^2)^2}\,\mathrm{d}(u_1^2) + \sqrt{\frac{u_1^2+b_1^2}{u_1^2+a_1^2}}\,\mathrm{d}u_1 =$$

$$= \left((a_1^2-b_1^2)u_1^2\frac{(u_1^2+a_1^2)^{-3/2}}{\sqrt{u_1^2+b_1^2}} + \sqrt{\frac{u_1^2+b_1^2}{u_1^2+a_1^2}}\right)\mathrm{d}u_1.$$

Стандартные преобразования дают

$$\mathrm{d}u = \frac{(a_1^2-b_1^2)u_1^2 + (u_1^2+a_1^2)(u_1^2+b_1^2)}{u_1^2+a_1^2}\,\frac{\mathrm{d}u_1}{\sqrt{(u_1^2+a_1^2)(u_1^2+b_1^2)}} =$$

$$= \frac{u_1^4+2a_1^2u_1^2+a_1^2b_1^2}{u_1^2+a_1^2}\,\frac{\mathrm{d}u_1}{\sqrt{(u_1^2+a_1^2)(u_1^2+b_1^2)}} = \text{правой части (2.4).} \qquad \square$$

В числителе дифференциала $\mathrm{d}u$ возник квадратный трёхчлен

$$u_1^4+2a_1^2u_1^2+a_1^2b_1^2$$

от переменной u_1^2. Его дискриминант оказывается квадратом:

$$a_1^4-a_1^2b_1^2 = a_1^2(a_1^2-b_1^2) = a_1^2\left(\left(\frac{a+b}{2}\right)^2-ab\right) = a_1^2\frac{(a-b)^2}{4},$$

поэтому трёхчлен раскладывается на множители.

Лемма 2. *Выполнено равенство*

$$u_1^4+2a_1^2u_1^2+a_1^2b_1^2 = (u_1^2+a_1b)(u_1^2+a_1a).$$

Доказательство. Действительно,

$$u_1^4+2a_1^2u_1^2+a_1^2b_1^2 = (u_1^2+a_1^2)^2 - a_1^2\cdot\frac{(a-b)^2}{4} =$$

$$= \left(u_1^2 + a_1^2 - a_1 \cdot \frac{a-b}{2}\right)\left(u_1^2 + a_1^2 + a_1 \cdot \frac{a-b}{2}\right) =$$
$$= \left(u_1^2 + a_1\left(a_1 - \frac{a-b}{2}\right)\right)\left(u_1^2 + a_1\left(a_1 + \frac{a-b}{2}\right)\right) = (u_1^2 + a_1 b)(u_1^2 + a_1 a). \quad \square$$

Учитывая полученное в лемме 2 разложение, перепишем связывающую дифференциалы формулу (2.4):

Лемма 1′. *Выполнено равенство*

$$\mathrm{d}u = \frac{(u_1^2 + aa_1)(u_1^2 + ba_1)}{u_1^2 + a_1^2} \frac{\mathrm{d}u_1}{\sqrt{(u_1^2 + a_1^2)(u_1^2 + b_1^2)}}. \tag{2.4′}$$

Теперь преобразуем подкоренные сомножители исходного дифференциала.

Лемма 3. *Имеют место равенства*

$$u^2 + a^2 = \frac{(u_1^2 + aa_1)^2}{u_1^2 + a_1^2}, \tag{2.5}$$
$$u^2 + b^2 = \frac{(u_1^2 + ba_1)^2}{u_1^2 + a_1^2}. \tag{2.6}$$

Доказательство. Преобразуем:

$$u^2 + a^2 =_{(2.2)} \frac{u_1^2 + b_1^2}{u_1^2 + a_1^2} \cdot u_1^2 + a^2 =$$
$$= \frac{(u_1^2 + b_1^2)u_1^2 + (u_1^2 + a_1^2)a^2}{u_1^2 + a_1^2} = \frac{u_1^4 + (a^2 + b_1^2)u_1^2 + a^2 a_1^2}{u_1^2 + a_1^2}. \tag{2.7}$$

Дискриминант числителя (как многочлена от u_1^2) равен

$$(a^2 + b_1^2)^2 - 4a^2 a_1^2 =_{(2.1)} (a^2 + ab)^2 - a^2(a+b)^2 = 0,$$

так что этот числитель — квадрат многочлена от u_1^2. Действительно, продолжая преобразования, получаем

$$u^2 + a^2 =_{(2.6)} \frac{\left(u_1^2 + \frac{a^2 + b_1^2}{2}\right)^2 + a^2 a_1^2 - \left(\frac{a^2 + b_1^2}{2}\right)^2}{u_1^2 + a_1^2} =$$
$$= \frac{\left(u_1^2 + \frac{a^2 + ab}{2}\right)^2 + a^2 a_1^2 - \left(\frac{a^2 + ab}{2}\right)^2}{u_1^2 + a_1^2} =_{(2.2)} = \frac{(u_1^2 + aa_1)^2}{u_1^2 + a_1^2}.$$

Аналогично

$$u^2 + b^2 =_{(2.2)} \frac{u_1^2 + b_1^2}{u_1^2 + a_1^2} \cdot u_1^2 + b^2 = \frac{(u_1^2 + b_1^2)u_1^2 + (u_1^2 + a_1^2)b^2}{u_1^2 + a_1^2} = \frac{u_1^4 + (b^2 + b_1^2)u_1^2 + b^2 a_1^2}{a_1^2 + u_1^2}. \tag{2.8}$$

Снова вычисляя дискриминант, получаем

$$(b^2+b_1^2)^2-4b^2a_1^2=_{(2.1)}(b^2+ab)^2-b^2(a+b)^2=0,$$

так что этот числитель — тоже квадрат многочлена от u_1^2. Продолжая, получаем

$$u^2+b^2=_{(2.8)}\frac{\left(u_1^2+\frac{b^2+b_1^2}{2}\right)^2+b^2a_1^2-\left(\frac{b^2+b_1^2}{2}\right)^2}{u_1^2+a_1^2}=$$

$$=\frac{\left(u_1^2+\frac{b^2+ab}{2}\right)^2+b^2a_1^2-\left(\frac{b^2+ab}{2}\right)^2}{u_1^2+a_1^2}=_{(2.2)}=\frac{(u_1^2+ba_1)^2}{u_1^2+a_1^2}.\qquad\square$$

Соединяя (2.5) и (2.6), получаем

$$\sqrt{(u^2+a^2)(u^2+b^2)}=\frac{(u_1^2+aa_1)(u_1^2+ba_1)}{u_1^2+a_1^2}.\tag{2.9}$$

Эта формула вместе с (2.4)′ доказывает наше предложение. □

§ 3. Алгебро-геометрический смысл подстановки Гаусса

Загадочная подстановка (2.2)

$$u=\sqrt{\frac{u_1^2+b_1^2}{u_1^2+a_1^2}}\,u_1\tag{3.1}$$

должна найти объяснение!

Все преобразования § 2 имеют ясный геометрический смысл; он не очень сложен, но довольно-таки далёк от школьной математики, поэтому объяснения получатся длинными.

3.0. Появление алгебраических кривых

Начнём с ещё одного шага от анализа к алгебре (точнее, к алгебраической геометрии): если раньше мы перешли от определённого интеграла

$$\int_{-\infty}^{\infty}\frac{\mathrm{d}u}{\sqrt{(u^2+a^2)(u^2+b^2)}}$$

к неопределённому

$$\int\frac{\mathrm{d}u}{\sqrt{(u^2+a^2)(u^2+b^2)}},$$

— а точнее, про интегралы почти забыли и работали с дифференциалами — то теперь давайте сосредоточимся на знаменателе подынтегрального выражения. Попробуем дать ему имя:

$$v := \sqrt{(u^2+a^2)(u^2+b^2)}. \tag{3.2}$$

Если мы хотим оставаться в рамках вещественной математики, то никаких трудностей в работе с этим новым действующим лицом не видно: подкоренное выражение всегда неотрицательно.

Но тут мы должны честно предупредить читателя, что понять происходящее, видя только вещественные числа, абсолютно невозможно и нам в скором времени придётся выйти в комплексную область, а там функция $z \mapsto \sqrt{z}$ определена плохо.

Выход из этого затруднения весьма прост: возведём соотношение (3.2) в квадрат, получив a, b, в координатах u, v с параметрами a, b уравнение *алгебраической кривой* которую мы назовём

$$\ddot{\mathbf{E}}_{a,b}\colon\ v^2 = (u^2+a^2)(u^2+b^2) \tag{3.3}$$

— это обозначение (в частности, две точки над именем кривой) скоро будет объяснено.

3.1. Изогения Гаусса

Теория кривых — богатый раздел алгебраической геометрии. Мы, однако, сразу обратимся к *отображениям* алгебраических кривых, причём весьма специального вида.

Наша ближайшая цель — установить связь между уравнением (3.3) и формулой (3.1). В координатах u_1, v_1 рассмотрим уравнение «другой» кривой

$$\ddot{\mathbf{E}}_{a_1,b_1}\colon\ v_1^2 = (u_1^2+a_1^2)(u_1^2+b_1^2). \tag{3.4}$$

Теперь мы готовы сформулировать результат, объясняющий происходящее.

Теорема-определение. *Определено отображение алгебраических кривых*

$$\ddot{\gamma}_{a,b}\colon\ \ddot{\mathbf{E}}_{a_1,b_1} \to \ddot{\mathbf{E}}_{a,b}\colon\ (u_1,v_1) \mapsto \left(\frac{u_1 v_1}{u_1^2+a_1^2}, \frac{u_1^4+2a_1^2u_1^2+a_1^2b_1^2}{u_1^2+a_1^2}\right).$$

Мы будем называть его изогенией Гаусса[5)].

Важное уточнение. Фигурирующее в теореме отображение не требует комментариев над полем действительных чисел $\mathbb{R}$: оно всюду опре-

[5)] Ни название, ни обозначение этого отображения не является общепринятым.

делено, поскольку при $a_1 \neq 0$ знаменатели *строго положительны*. Однако над произвольным полем — и прежде всего над особо интересующим нас полем комплексных чисел $\mathbb{C}$ — отображение $\ddot{\gamma}_{a,b}$ не определено в точках $(u_1 = \pm i a_1,\ v_1 = 0)$. Временно мы будем рассматривать изогению Гаусса вне этих точек, а впоследствии доопределим её.

Доказательство. Прямая проверка:

$$v^2 = \left(\frac{u_1^4 + 2a_1^2u_1^2 + a_1^2b_1^2}{u_1^2 + a_1^2}\right)^2 = \left(\frac{u_1^4 + 2\left(\frac{a+b}{2}\right)^2 u_1^2 + \left(\frac{a+b}{2}\right)^2 ab}{\left(\frac{a+b}{2}\right)^2 + u_1^2}\right)^2 \stackrel{?}{=}$$

$$\stackrel{?}{=} \left(\left(\frac{u_1v_1}{u_1^2 + a_1^2}\right)^2 + a^2\right)\left(\left(\frac{u_1v_1}{u_1^2 + a_1^2}\right)^2 + b^2\right)$$

или

$$\left(\frac{4u_1^4 + 2(a+b)^2u_1^2 + (a+b)^2ab}{4u_1^2 + (a+b)^2}\right)^2 \stackrel{?}{=} \left(\frac{u_1^2v_1^2}{(u_1^2 + a_1^2)^2} + a^2\right)\left(\frac{u_1^2v_1^2}{(u_1^2 + a_1^2)^2} + b^2\right).$$

Подставив в правую часть последнего равенства v_1^2 из уравнения (3.4), получим тождество, в которое входит только одна переменная u_1:

$$\left(\frac{4u_1^4 + 2(a+b)^2u_1^2 + (a+b)^2ab}{4u_1^2 + (a+b)^2}\right)^2 \stackrel{?}{=} \left(\frac{u_1^2v_1^2}{(u_1^2 + a_1^2)^2} + a^2\right)\left(\frac{u_1^2v_1^2}{(u_1^2 + a_1^2)^2} + b^2\right) =_{(3.4)}$$

$$=_{(3.4)} \left(\frac{(u_1^2 + a_1^2)(u_1^2 + b_1^2)u_1^2}{(u_1^2 + a_1^2)^2} + a^2\right)\left(\frac{(u_1^2 + a_1^2)(u_1^2 + b_1^2)u_1^2}{(u_1^2 + a_1^2)^2} + b^2\right) =$$

$$= \left(\frac{(u_1^2 + b_1^2)u_1^2}{u_1^2 + a_1^2} + a^2\right)\left(\frac{(u_1^2 + b_1^2)u_1^2}{u_1^2 + a_1^2} + b^2\right) =$$

$$= \frac{(u_1^2 + b_1^2)u_1^2 + a^2(u_1^2 + a_1^2)}{u_1^2 + a_1^2} \cdot \frac{(u_1^2 + b_1^2)u_1^2 + b^2(u_1^2 + a_1^2)}{u_1^2 + a_1^2} =$$

$$= \frac{\left(u_1^4 + (a^2 + b_1^2)u_1^2 + a^2a_1^2\right)\left(u_1^4 + (b^2 + b_1^2)u_1^2 + b^2a_1^2\right)}{(u_1^2 + a_1^2)^2}.$$

Для завершения доказательства остаётся воспользоваться тождествами

$$u_1^4 + (a^2 + b_1^2)u_1^2 + a^2a_1^2 = u_1^4 + (a^2 + ab)u_1^2 + a^2\frac{(a+b)^2}{4} = \left(u_1^2 + \frac{a(a+b)}{2}\right)^2$$

и

$$u_1^4 + (b^2 + b_1^2)u_1^2 + b^2a_1^2 = u_1^4 + (b^2 + ab)u_1^2 + b^2\frac{(a+b)^2}{4} = \left(u_1^2 + \frac{b(a+b)}{2}\right)^2. \quad \square$$

Теорема. *У каждой точки кривой* $\ddot{\mathbf{E}}_{a,b}$ *при изогении Гаусса* $\ddot{\gamma}_{a,b}$ *ровно два прообраза.*

Два доказательства. Сначала наметим «школьный» вариант. Надо исследовать систему уравнений

$$\begin{cases} u = \dfrac{u_1 v_1}{u_1^2 + a_1^2}, \\ v = \dfrac{u_1^4 + 2a_1^2 u_1^2 + a_1^2 b_1^2}{u_1^2 + a_1^2} \end{cases}$$

относительно неизвестных u_1, v_1. Второе уравнение является квадратным относительно u_1^2, так что при ненулевом дискриминанте у него ровно два решения. Величина v_1 определится первым уравнением при условии $u_1^2 + a_1^2 \neq 0$. Рассмотрение особых случаев предоставляется читателю.

Теперь проведём «взрослое» доказательство. Впрочем, главное соображение, на котором оно основано, вполне доступно школьнику:

Смена знаков обеих координат переводит кривые $\ddot{\mathbf{E}}_{a,b}$ *и* $\ddot{\mathbf{E}}_{a_1,b_1}$ *в себя*; *при этой смене знаков образ точки* $(u_1, v_1) \in \ddot{\mathbf{E}}_{a_1,b_1}$ *при изогении Гаусса не меняется*.

Иначе говоря, для любой точки $(u_1, v_1) \in \ddot{\mathbf{E}}_{a_1,b_1}$ верно, что

$$(-u_1, -v_1) \in \ddot{\mathbf{E}}_{a_1,b_1} \quad \text{и} \quad \ddot{\gamma}_{a,b}(u_1, v_1) = \ddot{\gamma}_{a,b}(-u_1, -v_1).$$

Теперь лемма мгновенно следует из теории Галуа. □

3.2. Эллиптические кривые

Мы собираемся разработать зрительные образы, связанные с кривыми $\ddot{\mathbf{E}}_{a_1,b_1}$ и $\ddot{\mathbf{E}}_{a,b}$; поскольку они принадлежат одному и тому же классу кривых, в дальнейшем мы будем говорить о $\ddot{\mathbf{E}}_{a,b}$.

Эти кривые называются *эллиптическими*, поскольку возникли при вычислении длины эллипса (см. введение).

Если эллипс задан уравнением

$$\frac{X^2}{A^2} + \frac{Y^2}{B^2} = 1,$$

то он допускает параметризацию $(X = A\cos t,\ Y = B\sin t)$ и длина дуги такого эллипса, которую можно символически выразить в виде

$$\int \sqrt{(\mathrm{d}X)^2 + (\mathrm{d}Y)^2},$$

задаётся интегралом

$$\int \sqrt{A^2 \sin^2 t + B^2 \cos^2 t}\, \mathrm{d}t.$$

После замены переменной $s := \sin t$ он превращается в

$$\int \sqrt{B^2 + (A^2 - B^2)s^2}\,\frac{\mathrm{d}s}{\sqrt{1-s^2}} = B \int \sqrt{1-k^2s^2}\,\frac{\mathrm{d}s}{\sqrt{1-s^2}},$$

где $k^2 = (B^2 - A^2)/B^2$. Последний интеграл с точностью до множителя традиционно переписывается в виде

$$\int \frac{(1-k^2s^2)\,\mathrm{d}s}{\sqrt{(1-s^2)(1-k^2s^2)}}$$

и распадается на два (они называются эллиптическими интегралами *первого* и *второго* рода). Один из них после очевидных преобразований (впрочем, невозможных в вещественной области) превращается в тот самый интеграл, которому посвящена настоящая статья.

Эллиптическая кривая

$$\ddot{\mathbf{E}}_k\colon\ v^2 = (1-u^2)(1-k^2u^2),$$

возникшая в результате этих вычислений (мы пользовались переменной s вместо u, чтобы не перепутать её с переменной в нашем основном интеграле) называется *квартикой Лежандра*.

Как мы видим, эллиптические кривые записываются (уже сотни лет) в разных формах; нам предстоит освоить ещё по крайней мере одну.

3.3. От квартик к кубикам

Некоторое время мы воздержимся от извлечения квадратных корней и от интегрирования, так что будем считать все переменные и коэффициенты многочленов принадлежащими произвольному *алгебраически замкнутому полю* (которое мы будем называть *основным полем*). Читатель, у которого это понятие вызывает затруднение, может считать, что мы работаем над полем комплексных чисел (только временно ограничиваемся четырьмя арифметическими операциями).

До сих пор мы работали с эллиптическими кривыми, задаваемыми уравнениями

$$v^2 = f_4(u), \tag{3.5}$$

где f_4 — многочлен 4-й степени; предположим, что он *не имеет кратных корней*. Рассмотренные нами многочлены даже были *чётны*, что облегчало рассмотрение симметрий (изогении определялись сменой знаков).

Следующий шаг может показаться техническим, но он весьма важен. Мы собираемся перейти от многочленов степени 4 к многочленам степени 3.

Пользуясь алгебраической замкнутостью основного поля, разложим правую часть уравнения (3.5) на множители:

$$v^2 = (u-\alpha_1)(u-\alpha_2)(u-\alpha_3)(u-\alpha_4); \tag{3.6}$$

согласно нашему предположению, все числа $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ *разные*. Применим к уравнению (3.6) любое дробно-линейное преобразование

$$u \to \frac{au+b}{cu+d},$$

переводящее какой-либо из корней $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ в *бесконечно удалённую точку*; нам удобно будет выбрать такое, что другой корень переведётся в 0. Иначе говоря, введём новую координату (где $m \neq 0$ остаётся в нашем распоряжении)

$$x := m\frac{u-\alpha_3}{u-\alpha_4}. \tag{3.7}$$

Исходная координата u легко выразится через неё:

$$u = \frac{\alpha_4 x - m\alpha_3}{x-m}. \tag{3.8}$$

Когда мы подставим соотношение (3.8) в уравнение (3.6), это уравнение (временно) утратит полиномиальность, превратившись в

$$v^2 = m(\alpha_3-\alpha_4)^2\frac{x((\alpha_1-\alpha_4)x-m\alpha_1+m\alpha_3)((\alpha_2-\alpha_4)x-m\alpha_2+m\alpha_3)}{(x-m)^4}, \tag{3.9}$$

или

$$v^2 = m\frac{\beta_{34}^2 x(\beta_{14}x-m\beta_{12})(\beta_{24}x-m\beta_{23})}{(x-m)^4}, \tag{3.10}$$

где для краткости записи введены величины $\beta_{ij} = \alpha_i - \alpha_j$; отметим, что из несовпадения α_i следует

$$\beta_{34}^2\beta_{14}\beta_{24} \neq 0, \tag{3.11}$$

т. е. в числителе правой части уравнения (3.10) стоит именно кубический многочлен. Полиномиальность уравнения рассматриваемой кривой восстановится, если ввести новую координату y соотношением

$$v = \frac{\lambda y}{(x-m)^2}, \tag{3.12}$$

где λ — константа, которую мы сейчас подберём. Подставив (3.12) в (3.10) и умножив полученное уравнение на $(x-m)^4$, придём к уравнению

$$\lambda^2 y^2 = m\beta_{34}^2 x(\beta_{14}x-m\beta_{12})(\beta_{24}x-m\beta_{23}). \tag{3.13}$$

Подобрав подходящую константу λ (здесь мы очередной раз пользуемся алгебраической замкнутостью основного поля), мы придём к уравнению

$$y^2 = x(x-\gamma_1)(x-\gamma_2) \tag{3.14},$$

в которое, таким образом, преобразовалось исходное уравнение (3.6). Легко проверить, что преобразование координат (x, y) в координаты (u, v), задаваемое уравнениями (3.8) и (3.12), обратимо. Остаётся (временное) затруднение, связанное с тем, что и это преобразование, и обратное ему не определены в конечном числе точек.

Обратимые преобразования, в обоих направлениях заданные рациональными функциями, называются *бирациональными*. Мы по существу показали, что *любая кривая, заданная уравнением* $v^2 = f_4(u)$, *бирационально эквивалентна кривой, заданной уравнением* $y^2 = f_3(x)$, *где* f_3 *и* f_4 — *многочлены* степени 3 и 4, не имеющие кратных корней. Проанализируем более детально эту бирациональную эквивалентность — обычно она называется *бирациональным изоморфизмом* — для наших кривых $\ddot{\mathbf{E}}_{a,b}$.

3.4. Преобразование квартики $\ddot{\mathbf{E}}_{a,b}$ в кубику $\dot{\mathbf{E}}_{a,b}$

В случае кривой $\ddot{\mathbf{E}}_{a,b}$, заданной уравнением (3.3), имеем для уравнения в форме (3.5):

$$f_4(u) = (u^2 + a^2)(u^2 + b^2).$$

Начиная с этого места будем пропускать детали вычислений и приводить лишь ответы. Очень усердный читатель восстановит вычисления вручную, идя вслед за Гауссом и его толкователями XIX века; автор, однако, настоятельно рекомендует использовать современные системы компьютерной алгебры.

Проведённое в пункте 3.3 преобразование квартики $\ddot{\mathbf{E}}_{a,b}$ в кубику $\dot{\mathbf{E}}_{a,b}$ использует *мнимую единицу* i; можно считать, что мы стали работать над полем комплексных чисел, а можно — что мы выбрали (один из двух) элементов основного поля[6)], удовлетворяющих равенству $\mathrm{i}^2 = -1$.

Наше преобразование осуществляется формулами

$$x = -\frac{b(a+b)}{2}\,\frac{u-\mathrm{i}a}{u-\mathrm{i}b}, \quad y = \frac{\mathrm{i}b(a^2-b^2)}{4}\,\frac{v}{(u-\mathrm{i}b)^2}. \tag{3.15}$$

Мы готовы перейти к кубике, которая сыграет основную роль в понимании изогении Гаусса.

[6)] Здесь надо предположить, что характеристика основного поля отлична от 2, т. е. $1+1 \neq 0$.

Теорема-обозначение. *Преобразование* (3.15) *переводит кривую* $\ddot{\mathbf{E}}_{a,b}$, *заданную, напомним, уравнением*

$$v^2 = (u^2 + a^2)(u^2 + b^2),$$

в кривую[7] $\dot{\mathbf{E}}_{a,b}$, *заданную уравнением*

$$y^2 = x(x + ab)\Big(x + \frac{(a+b)^2}{4}\Big). \tag{3.16}$$

Замечания. (1) Уравнение кривой $\dot{\mathbf{E}}_{a,b}$ легко запомнить: ненулевые корни его правой части — взятые с обратным знаком *квадраты среднего арифметического и среднего геометрического* от параметров кривой.

(2) Преобразование (3.15) не имеет смысла при $a = \pm b$: выражения для координат превращаются в константы. Поэтому над любым полем мы предполагаем $a \neq \pm b$.

(3) Преобразование не имеет смысла в характеристике 2, так что мы исключаем этот случай из рассмотрения.

(4) Преобразование *кажется* не имеющим смысла при $u = \mathrm{i}b$, а обратное ему преобразование (выписать которое предлагается читателю) — при $x = -b(a+b)/2$. Однако это впечатление ложно, и мы сейчас с этим будем разбираться.

(5) Многочлен в правой части (3.16) при $a \neq \pm b$ не имеет кратных корней.

3.5. Проективизация

До сих пор нам не требовалось обозначение для *основного поля*: мы просто проводили алгебраические операции над константами и координатами. *Теперь мы обозначим это поле традиционной буквой*[8] $\Bbbk$. Напомним, что мы считаем поле $\Bbbk$ произвольным *алгебраически замкнутым* и предполагаем, что $\mathrm{char}(\Bbbk) \neq 2$. Читатель в очередной раз может считать, что $\Bbbk = \mathbb{C}$, но это не облегчает понимание обсуждаемых алгебро-геометрических конструкций.

До сих пор мы считали рассматриваемые кривые подмножествами плоскости $\Bbbk \times \Bbbk$, которую теперь будем называть *аффинной* и обозначать

$$\mathbf{A}_2(\Bbbk) := \Bbbk \times \Bbbk;$$

её подмножества, задаваемые полиномиальными уравнениями, мы теперь будем называть *аффинными кривыми*.

---

[7] Обратите внимание на количество точек над **E**!

[8] От немецкого *Körper* — тело. Термин был предложен Р. Дедекиндом, чтобы подчеркнуть *целостность* (числовых) множеств, замкнутых относительно арифметических операций.

Аффинную плоскость будем считать подмножеством *проективной*: $\mathbf{A}_2(\Bbbk) \hookrightarrow \mathbf{P}_2(\Bbbk)$. Проективная плоскость $\mathbf{P}_2(\Bbbk)$ — это следующее фактормножество:

$$\mathbf{P}_2(\Bbbk) := \frac{(\Bbbk \times \Bbbk \times \Bbbk) \setminus \{(0,0,0)\}}{\Bbbk^\times},$$

где $\Bbbk^\times$ — мультипликативная группа поля, действующая на тройки элементов поля покомпонентным умножением. Будем обозначать через $(x:y:z)$ орбиту элемента $(x,y,z) \in \Bbbk \times \Bbbk \times \Bbbk$; тогда вложение $\mathbf{A}_2(\Bbbk) \hookrightarrow \mathbf{P}_2(\Bbbk)$ задаётся формулой

$$(x,y) \mapsto (x:y:1).$$

Мы собираемся перейти от аффинных кривых к проективным. Для кривых $y^2 = f_3(x)$, где f_3 — многочлен без кратных корней, это сделать совсем легко; разберём случай

$$\dot{\mathbf{E}}_{\gamma_1,\gamma_2}\colon\ y^2 = x(x-\gamma_1)(x-\gamma_2), \tag{3.17}$$

где $\gamma_1, \gamma_2 \in \dot{\Bbbk} := \Bbbk \setminus \{0\}$ и $\gamma_1 \neq \gamma_2$. Согласно (3.16) и (3.17) кривая $\dot{\mathbf{E}}_{\gamma_1,\gamma_2}$ превращается в $\dot{\mathbf{E}}_{a,b}$ при

$$\gamma_1 = -ab, \gamma_2 = -\frac{(a+b)^2}{4}. \tag{3.18}$$

В проективной плоскости $\mathbf{P}_2(\Bbbk)$ мы будем пользоваться *однородными* координатами $(x:y:z)$, и в них аффинная плоскость $\mathbf{A}_2(\Bbbk) \subset \mathbf{P}_2(\Bbbk)$ определяется условием $z \neq 0$. При этом условии имеет место равенство

$$(x:y:z) = \left(\frac{x}{z} : \frac{y}{z} : 1\right)$$

и уравнение (3.17) можно переписать в виде

$$\left(\frac{y}{z}\right)^2 = \frac{x}{z}\left(\frac{x}{z} - \gamma_1\right)\left(\frac{x}{z} - \gamma_2\right). \tag{3.19}$$

Умножив это уравнение на z^3, мы получим уравнение кривой в $\mathbf{P}_2(\Bbbk)$, которую обозначим

$$\mathbf{E}_{\gamma_1,\gamma_2}\colon y^2 z = x(x-\gamma_1 z)(x-\gamma_2 z) \tag{3.20}$$

(обратите внимание на отсутствие точки над **E**!). Из определения вложения аффинной плоскости в проективную и из соотношения между аффинными координатами (x,y) и однородными $(x:y:z)$ следует, что

$$\dot{\mathbf{E}}_{\gamma_1,\gamma_2} \subset \mathbf{E}_{\gamma_1,\gamma_2}$$

и что в однородных координатах

$$\dot{\mathbf{E}}_{\gamma_1,\gamma_2} = \{P \in \mathbf{E}_{\gamma_1,\gamma_2} \mid z(P) \neq 0\}. \tag{3.21}$$

Подчеркнём, что хотя значение $z(P)$ координаты z в точке P определено лишь с точностью до ненулевого множителя, обращение $z(P)$ в нуль определено корректно.

Остаётся определить разность $\mathbf{E}_{\gamma_1,\gamma_2} \setminus \dot{\mathbf{E}}_{\gamma_1,\gamma_2}$. Для этого надо определить пересечение кривой $\mathbf{E}_{\gamma_1,\gamma_2}$ с так называемой *бесконечно удалённой прямой*

$$\ell_\infty\colon\ z = 0. \tag{3.22}$$

Это просто: подстановка (3.22) в (3.20) даёт

$$0 = x^3, \tag{3.23}$$

т. е. на пересечении $\ell_\infty \bigcap \mathbf{E}_t$ выполнены два равенства: $x = 0$ и $z = 0$, а это возможно лишь в одной точке

$$\underline{\infty} := (0 : 1 : 0) \in \mathbf{P}_2(\Bbbk).$$

Мы установили, что

$$\ell_\infty \bigcap \mathbf{E}_t = \{\underline{\infty}\}$$

(на самом деле формула (3.23) показывает, что это пересечение *трёхкратно*, но мы в настоящей статье не касаемся теории кратностей и лишь упомянем, что ℓ_∞ — одна из *прямых перегиба* кривой $\mathbf{E}_{\gamma_1,\gamma_2}$). Кроме того, из проведённых рассмотрений следует, что

$$\dot{\mathbf{E}}_{\gamma_1,\gamma_2} = \mathbf{E}_{\gamma_1,\gamma_2} \setminus \{\underline{\infty}\},$$

и мы наконец можем объяснить наше обозначение $\dot{\mathbf{E}}$: точка над обозначением *проективной* кривой $\mathbf{E}$ означает *прокол*!

Итак, нам удалось реализовать *аффинную* кубику $\dot{\mathbf{E}}_{a,b}$ как *проективную* кубику $\mathbf{E}_{a,b}$ с одним проколом. Эта проективная кубика *гладка*; читателю предлагается сформулировать это свойство (или найти его в любом учебнике алгебраической геометрии); первоочередная задача — понять *топологию* этих кубик над $\mathbb{C}$; это — одна из промежуточных целей нашего выхода из аффинной плоскости в проективную.

Однако наша окончательная цель — понимание бесконечной цепочки изогений $\mathbf{E}_{a_{n+1},\,b_{n+1}} \to \mathbf{E}_{a_n,b_n}$ — потребует некоторых дополнительных средств.

3.6. Пополнение квартики $\ddot{\mathbf{E}}_{a,b}$

Нам встречались и обозначения $\ddot{\mathbf{E}}$ с *двумя* точками над обозначением кривой; разберём основной для нас случай кривых $\ddot{\mathbf{E}}_{a,b}$. Здесь проективизация не даёт желаемого результата, поскольку «на бесконечности»

в $\mathbf{P}_2(\mathbb{k})$ кривая оказывается *особой*, т. е. *не*гладкой (читатель может придать точный смысл этому утверждению и проверить его — или же принять на веру). Приходится идти другим путём.

Начнём с интуитивного соображения. Если u стремится к бесконечности, то константы в уравнении $v^2 = (u^2 + a^2)(u^2 + b^2)$ пренебрежимо малы по сравнению с u и выполняется приближённое уравнение $v^2 \approx u^4$, обладающее двумя приближёнными решениями $v \approx \pm u^2$. Это позволяет заподозрить, что существует *полная*[9] кривая $\mathbf{E}_{a,b}$ (скоро мы прокомментируем повторное использование этого обозначения) и две бесконечно удалённые точки на ней $\underline{\infty}^{\pm} \in \mathbf{E}_{a,b}$, такие, что

$$\ddot{\mathbf{E}}_{a,b} = \mathbf{E}_{a,b} \setminus \{\underline{\infty}^{+}, \underline{\infty}^{-}\}.$$

На интуитивном уровне мы объяснили использование обозначения $\ddot{\mathbf{E}}$ как имеющего смысл «два прокола».

Чтобы придать этому построению точный смысл, введём на кривой $\ddot{\mathbf{E}}_{a,b}$ новые *локальные* (т. е. имеющие смысл вне конечного множества) координаты

$$U := \frac{1}{u^2}, \quad V := \frac{v}{u^2}, \tag{3.24}$$

связанные соотношением

$$V^2 = (a^2U^2 + 1)(b^2U^2 + 1). \tag{3.25}$$

Добавим ещё две точки $O^{\pm} \in \ddot{\mathbf{E}}_{a,b}$ условиями

$$u(O^{\pm}) = 0, \quad v(O^{\pm}) = \pm ab. \tag{3.26}.$$

Координаты (U, V) определены лишь на множестве $\ddot{\mathbf{E}}_{a,b} \setminus \{O^{\pm}\}$; зато, временно введя аффинную кривую $'\ddot{\mathbf{E}}_{a,b}$, определённую уравнением (3.25), мы обнаружим на ней две точки, которые выше определялись неформально:

$$U(\underline{\infty}^{\pm}) = 0, \quad V(\underline{\infty}^{\pm}) = \pm 1. \tag{3.27}$$

Теперь мы можем определить *абстрактную* кривую

$$\mathbf{E}_{a,b} := \ddot{\mathbf{E}}_{a,b} \bigcup {}'\ddot{\mathbf{E}}_{a,b},$$

где на каждой из объединяемых аффинных кривых подразумеваются свои локальные координаты, связанные на пересечении соотношениями (3.24).

По аналогии с вложением кубики в проективную плоскость может возникнуть желание вложить в какое-нибудь проективное простран-

[9] Для кривой это означает, что она определена всюду вне конечного множества значений координат.

ство (или в произведение проективных пространств) полученное объединение аффинных кривых. Сделать это можно, хотя очевидных решений, видимо, нет. Мы, однако, заниматься такими вложениями не будем: математики переходят от вложенных многообразий к абстрактным, начиная с Гаусса и Римана в XIX веке — когда, впрочем, *алгебраические* многообразия понимались в основном как вложенные в аффинные или проективные пространства; лишь в XX веке абстрактные многообразия стали основными. В настоящее время издано, а также выложено в интернет огромное количество учебников и лекционных курсов по современной алгебраической геометрии; русскоязычному читателю естественно порекомендовать [7], где он найдёт и определение *полного* многообразия.

Мы позволили себе отождествить рассматриваемое объединение двух аффинных квартик с проективной кубикой не только потому, что они *изоморфны*. Читателю предлагается проверить, что формула (3.15) (относительно которой отмечалось, что соответствующее отображение только кажется неопределённым в некоторых точках) на самом деле определяет изоморфизм пополненных кривых $\mathbf{E}_{ab} \xrightarrow{\simeq} \mathbf{E}_{ab}$. Главная причина заключается в том, что мы становимся на упомянутую только что современную точку зрения: считаем, что речь идёт об *одной и той же* кривой, только *реализованной* по-разному (в одном случае вложение в проективное пространство предъявлено, в другом — нет).

3.7. Квартика или кубика?

Каждая из реализаций имеет свои достоинства.

(а) Квартика $\ddot{\mathbf{E}}_{a,b}$ обладает редким для эллиптических кривых свойством: на ней сразу видны целых восемь точек! Помимо уже введённых четырёх, бросаются в глаза точки с u-абсциссами $\pm \mathrm{i}a$ и $\pm \mathrm{i}b$; дадим этим новым точкам легко запоминающиеся имена $A^{\pm}$ и $B^{\pm}$. Полный список выделенных точек и их координат представлен в табл. 1.

Здесь (x, y)-координаты точек вычисляются по формулам (3.15). На кубике $\mathbf{E}_{a,b}$ точки $O^{\pm}$ и $\underline{\infty}^{\pm}$ в глаза не бросаются.

(б) На квартике действует *четверная группа Клейна* $\mathsf{C}_2 \times \mathsf{C}_2$, состоящая, помимо тождественного отображения, из инволюций $(u, v) \mapsto (-u, v)$, $(u, v) \mapsto (u, -v)$ и $(u, v) \mapsto (-u, -v)$. Отметим, что все они сохраняют наши 8 точек. Наиболее важна для нашего основного сюжета инволюция

$$\iota \colon (u, v) \mapsto (-u, -v),$$

не имеющая неподвижных точек.

Таблица 1

Точка	x	y	u	v	U	V
B^+	∞	∞	$\mathrm{i}b$	0	$-\frac{\mathrm{i}}{b}$	0
B^-	$-\left(\frac{a+b}{2}\right)^2$	0	$-\mathrm{i}b$	0	$\frac{\mathrm{i}}{b}$	0
A^-	$-ab$	0	$-\mathrm{i}a$	0	$\frac{\mathrm{i}}{a}$	0
A^+	0	0	$\mathrm{i}a$	0	$-\frac{\mathrm{i}}{a}$	0
O^+	$-\frac{a(a+b)}{2}$	$\frac{-\mathrm{i}a(a^2-b^2)}{4}$	0	ab	∞	∞
O^-	$-\frac{a(a+b)}{2}$	$\frac{\mathrm{i}a(a^2-b^2)}{4}$	0	$-ab$	∞	∞
$\underline{\infty}^+$	$-\frac{b(a+b)}{2}$	$\frac{\mathrm{i}b(a^2-b^2)}{4}$	∞	∞	0	1
$\underline{\infty}^-$	$-\frac{b(a+b)}{2}$	$-\frac{\mathrm{i}b(a^2-b^2)}{4}$	∞	∞	0	-1

(в) Кубика $\dot{\mathbf{E}}_{a,b}$ — гладкая аффинная кривая, дополненная в проективной плоскости единственной точкой B^+ (признаемся — *случайно* выбранной из четырёх точек $A^\pm, B^\pm$), тоже гладкой на проективном замыкании $\mathbf{E}_{a,b} \subset P_2(\Bbbk)$.

(г) На любой кубике $\mathbf{E}$ имеется структура абелевой группы $(\mathbf{E}, \oplus)$. Эта структура однозначно определяется правилом для $P, Q, R \in \mathbf{E}$:

$$P \oplus Q \oplus R = 0_{\mathbf{E}} \iff P, Q \text{ и } R \text{ коллинеарны}^{10)}$$

и правилом выбора нейтрального элемента $0_{\mathbf{E}} =: O \in \mathbf{E}$. Эта классическая теория достаточно сложна, см. введение в неё, например, в [6]. В случае кубики, заданной уравнением $y^2 = f_3(x)$, в качестве нейтрального элемента традиционно выбирают «бесконечно удалённую» точку $O \in \mathbf{E} \setminus \dot{\mathbf{E}}$. Для нашей кривой мы в соответствии с этой традицией произведём переименование

$$B^+ =: O \in \mathbf{E}_{a,b}.$$

Школьной математики (теорем Виета…) достаточно, чтобы выписать сложение $\oplus$ формулами в координатах. Однако формулы получаются до-

[10)] В случае совпадения двух точек проходящая через них прямая заменяется на *касательную* к $\mathbf{E}$, а в случае совпадения трёх — на её *прямую перегиба*.

Таблица 2

Сложение $P \oplus Q$

Q \ P	O	B^-	A^-	A^+	O^+	O^-	$\underline{\infty}^+$	$\underline{\infty}^-$
O	O	B^-	A^-	A^+	O^+	O^-	$\underline{\infty}^+$	$\underline{\infty}^-$
B^-	B^-	O	A^+	A^-	O^-	O^+	$\underline{\infty}^-$	$\underline{\infty}^+$
A^-	A^-	A^+	O	B^-	$\underline{\infty}^-$	$\underline{\infty}^+$	O^-	O^+
A^+	A^+	A^-	B^-	O	$\underline{\infty}^+$	$\underline{\infty}^-$	O^+	O^-
O^+	O^+	O^-	$\underline{\infty}^-$	$\underline{\infty}^+$	B^-	O	A^-	A^+
O^-	O^-	O^+	$\underline{\infty}^+$	$\underline{\infty}^-$	O	B^-	A^+	A^-
$\underline{\infty}^+$	$\underline{\infty}^+$	$\underline{\infty}^-$	O^-	O^+	A^-	A^+	B^-	O
$\underline{\infty}^-$	$\underline{\infty}^-$	$\underline{\infty}^+$	O^+	O^-	A^+	A^-	O	B^-

статочно длинные[11)], и мы их не приводим. Тем не менее настоятельно рекомендуем читателю воспользоваться имеющимися техническими средствами, чтобы получать эти формулы и «нажатием кнопки» проверять наши утверждения, которые приводятся без обоснования.

Восемь выделенных нами (с помощью *квартики*) точек образуют *подгруппу*!

Если посмотреть на левую верхнюю четверть табл. 2, то видно, что множество $\{O, B^-, A^\pm\}$ составляет 4-элементную группу точек *порядка* 2, а правая нижняя четверть с учётом левой верхней показывает, что все восемь — точки *порядка* 4. (Для поля $\mathbb{C}$ скоро станет очевидно, что наша 8-элементная группа — половина 16-элементной группы точек порядка 4.)

3.8. Групповая интерпретация изогении Гаусса

Теперь пора вспомнить, что нас интересует не одна кривая $\mathbf{E}_{a,b}$, а изогения Гаусса

$$\mathbf{E}_{a_1,\,b_1} \xrightarrow{\gamma_{a,b}} \mathbf{E}_{a,b}$$

или, ещё лучше, — последовательность изогений

$$\dots \xrightarrow{\gamma_{a_2,\,b_2}} \mathbf{E}_{a_2,\,b_2} \xrightarrow{\gamma_{a_1,\,b_1}} \mathbf{E}_{a_1,\,b_1} \xrightarrow{\gamma_{a_0,\,b_0}} \mathbf{E}_{a_0,\,b_0},$$

— мы просто добавляем индексы $n = 0, 1, 2, \dots$ к введённым ранее обозначениям параметров, координат, точек, кривых и морфизмов. Предполага-

[11)] Например, прямая проверка *ассоциативности* операции $\oplus$ весьма громоздка, и вместо неё обычно используются нетривиальные геометрические конструкции — см. [6].

ется, что параметры $\{a_n, b_n\}$ связаны соотношениями $a_{n+1} = (a_n + b_n)/2$ и $b_{n+a}^2 = a_n b_n$ (поскольку мы продолжаем работать над почти произвольным полем, мы нехитрым трюком избежали извлечения корней[12]). Прямые вычисления показывают, что каждая изогения $\gamma_{a_n, b_n} : \mathbf{E}_{a_{n+1}, b_{n+1}} \to \mathbf{E}_{a_n, b_n}$ преобразует координаты кубики следующим образом:

$$x_n = \frac{4x_{n+1}(x_{n+1} + a_{n+1}b_{n+1})}{4x_{n+1} + (a_{n+1} + b_{n+1})^2},$$
$$y_n = -\frac{4y_{n+1}(2x_{n+1} + a_{n+1}^2 + a_{n+1}b_{n+1})(2x_{n+1} + b_{n+1}^2 + a_{n+1}b_{n+1})}{(4x_{n+1} + (a_{n+1} + b_{n+1})^2)^2}.$$

Небольшой прогресс по сравнению с изогениями в координатах квартик, которые мы взялись прояснить! Формулы стали только длиннее…

Но здесь нас и выручит групповая структура на кубиках. Подставляя в эти ужасные формулы координаты наших восьми точек, мы можем убедиться, что изогении Гаусса ведут себя очень просто:

$$B_{n+1}^{\pm} \mapsto B_n^{+}, \quad A_{n+1}^{\pm} \mapsto A_n^{+}, \quad O_{n+1}^{\pm} \mapsto B_n^{-}, \quad \underline{\infty}_{n+1}^{\pm} \mapsto A_n^{-}.$$

Оказывается, они обладают своеобразным свойством «чётности»: в очевидных обозначениях $\gamma_{a_n, b_n}(X^+) = \gamma_{a_n, b_n}(X^-)$. Это свойство следует сопоставить с не сразу бросающимся в глаза свойством приведённой выше таблицы сложения 8-элементной группы:

прибавление точки B^- «меняет знак»!

Вспомним теперь инволюции без неподвижных точек, которые здесь естественно обозначить ι_n и которые меняют знаки в прямом смысле слова (т. е. действуют по формуле $(u_n, v_n) \mapsto (-u_n, -v_n)$), и отождествить её и отождествим их с прибавлением B^- (это, как и многое другое, предоставляется читателю). Тогда обсуждаемый факт можно сформулировать более традиционно:

$$\gamma_{a_n, b_n} \circ \iota_n = \gamma_{a_n, b_n},$$

или[13]

$$\forall\, P \in \mathbf{E}_{a_n, b_n} \quad [\gamma_{a_n, b_n}(P \oplus B_n^-) = \gamma_{a_n, b_n}(P)].$$

Мы вплотную подошли к разгадке тайны изогении Гаусса. Приведённые формулировки по существу равносильны следующему результату:

[12] Гаусс работал также со случаем $a_n, b_n \in \mathbb{C}$, и преодоление неоднозначности извлечения корней из комплексных чисел привело его к замечательным теориям, сильно опередившим его время, — см. [2].

[13] Проверив прямым вычислением или выведя из каких-нибудь общих теорем, что такое соотношение достаточно проверить лишь на нескольких точках.

Теорема. *Каждая изогения Гаусса представляет собой факторизацию по двухэлементной группе*

$$\gamma_{a_n,b_n}\colon\ \mathbf{E}_{a_{n+1},\,b_{n+1}} \to \frac{\mathbf{E}_{a_{n+1},\,b_{n+1}}}{\{O_{n+1},\,B^-_{n+1}\}} \cong \mathbf{E}_{a_n,b_n}. \qquad \square$$

Краткая формулировка найдена. Осталось придать ей наглядность.

§ 4. Трансцендентный смысл изогении Гаусса

В этом последнем параграфе основным полем является $\Bbbk = \mathbb{C}$. Мы чуть-чуть коснёмся большой классической теории, взяв из неё лишь то, что нужно для наших целей.

4.0. Решётки в $\mathbb{C}$ и факторы по ним

Будем называть *решёткой* любую дискретную подгруппу ранга 2 аддитивной группы комплексных чисел. Иначе говоря, дискретное подмножество $\Lambda \subset \mathbb{C}$ — решётка, если $0 \in \Lambda$, $\Lambda + \Lambda \subset \Lambda$, $-\Lambda = \Lambda$ и имеет место изоморфизм групп $(\Lambda, +) \simeq \mathbb{Z} \times \mathbb{Z}$.

Мы введём обозначение $\dot{\mathbb{C}} := \mathbb{C} \setminus \{0\}$ и для решётки Λ соответственно $\dot{\Lambda} = \Lambda \setminus \{0\}$. Для $k \in \dot{\mathbb{C}}$ решётки Λ и $k\Lambda$ называются *подобными*. Любая решётка представима (не единственным образом) в виде $\Lambda = \mathbb{Z}\lambda_1 + \mathbb{Z}\lambda_2$, где $\lambda_1, \lambda_2 \in \dot{\mathbb{C}}$ и $\lambda_1/\lambda_2 \notin \mathbb{R}$. Из класса подобных решёток принято выбирать представитель вида $\Lambda = \mathbb{Z}\tau + \mathbb{Z}$, где $\operatorname{Im}(\tau) > 0$. Нам предстоит работать с *прямоугольными* решётками такого вида, где τ будет чисто мнимым числом. В частности, исходным примером для Гаусса была *квадратная решётка* с $\tau = \mathrm{i}$.

Фактор комплексной плоскости по решётке гомеоморфен тору. Этот факт особенно несомненен в случае прямоугольных решёток (которыми мы только и будем заниматься) в традиционной нормировке: *фундаментальная область* такой решётки — прямоугольник, и сдвиг по мнимой образующей решётки $z \mapsto z + \tau$ отождествляет горизонтальные стороны прямоугольника, а сдвиг по вещественной образующей $z \mapsto z + 1$ отождествляет вертикальные.

Особенно важно осмыслить не только фактор, но и *морфизм факторизации*

$$\mathbb{C} \to \frac{\mathbb{C}}{\Lambda}.$$

Предлагается воспринимать его как комплексный аналог морфизма

$$\mathbb{R} \to \frac{\mathbb{R}}{2\pi\mathbb{Z}},$$

в котором тор заменяется на (тригонометрическую) окружность, а сам морфизм осуществляется парой функций (cos, sin). Об аналоге этой пары функций мы сейчас бегло расскажем[14].

4.1. $\wp$-функция Вейерштрасса

Подробнее с материалом этого пункта можно познакомиться по книге [4].

Для любого $k \in \{2, 3, 4, \ldots\}$ введём *ряды Эйзенштейна* — функции решёток

$$G_k(\Lambda) := \sum_{\lambda \in \dot{\Lambda}} \frac{1}{\lambda^{2k}} \quad \text{и} \quad g_2(\Lambda) := 60G_2(\Lambda), \quad g_3(\Lambda) := 140G_3(\Lambda).$$

Главное действующее лицо — мероморфная[15] *$\wp$-функция*

$$\wp_\Lambda(z) := \frac{1}{z^2} + \sum_{\lambda \in \dot{\Lambda}} \Big(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \Big).$$

Её Λ-периодичность не совсем очевидна, но следует из её чётности и очевидной Λ-периодичности её производной

$$\wp'_\Lambda(z) = -2 \sum_{\lambda \in \Lambda} \frac{1}{(z-\lambda)^3}.$$

(аналогичная сумма для самой $\wp$-функции расходилась бы).

Оказывается, $\wp$-функция для любой решётки Λ удовлетворяет дифференциальному уравнению (тоже Вейерштрасса)

$$(\wp'_\Lambda(z))^2 = 4\wp_\Lambda(z)^3 - g_2(\Lambda)\wp_\Lambda(z) - g_3(\Lambda),$$

а это, с точностью до простых аффинных преобразований, — уравнение рассмотренных нами кубик, так что пара функций $(\wp_\Lambda, \wp'_\Lambda)$ действительно аналогична паре (cos, sin).

Вернёмся к нашим кубикам $\dot{\mathbf{E}}_{a,b}$, которые заданы, напомним, уравнением

$$y^2 = x(x+ab)\Big(x + \Big(\frac{a+b}{2}\Big)^2\Big),$$

где $a \geqslant b > 0$ согласно (2.0). Случай $a = b$ тривиален. При $a > b > 0$ получаем $-\big(\frac{a+b}{2}\big)^2 < -ab < 0$. Можно показать, что возникают два *периода*:

$$\text{вещественный} \quad \int\limits_{-\left(\frac{a+b}{2}\right)^2}^{-ab} \frac{dx}{y} \qquad \text{и} \qquad \text{мнимый} \quad i\int\limits_{-ab}^{0} \frac{dx}{\sqrt{-y^2}}$$

[14] Главное отличие заключается в том, что решётка $\mathbb{Z} \subset \mathbb{R}$ единственна, а классы подобия решёток $\Lambda \subset \mathbb{C}$ зависят от одного комплексного параметра.

[15] А именно, имеющая полюсы второго порядка в точках решётки Λ.

(при $b \to a \neq 0$ вещественный период, очевидно, стремится к нулю, а мнимый — к ненулевому мнимому числу). Решётка Λ, соответствующая кривой $\mathbf{E}_{a,b}$, порождена этими периодами.

Аффинные преобразования, приводящие используемые нами уравнения кривых $\dot{\mathbf{E}}_{a,b}$ к требуемому виду

$$Y^2 = 4X^3 - g_2 X - g_3,$$

достаточно громоздки, и мы ограничимся случаем квадратной решётки, в котором $a = a_0 = \sqrt{2}+1$, $b = b_0 = \sqrt{2}-1$, а уравнения кривой имеют вид

$$y^2 = x(x+1)(x+2), \quad v^2 = u^4 + 6u^2 + 1.$$

Здесь из соображений симметрии $G_3 = g_3 = 0$, а

$$G_2 = \sum_{(m,n)\in(\mathbb{Z}\times\mathbb{Z})\setminus\{(0,0)\}} \frac{1}{(m+n\mathrm{i})^4} = \frac{\varpi^4}{15} \approx 3{,}151,$$

$$g_2 = 60G_2 = 4\varpi^4 \approx 189{,}070.$$

Требуемое преобразование имеет вид

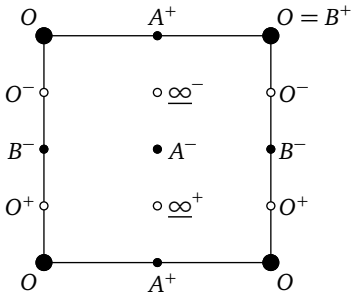
$$X := \frac{x}{\varpi^2} - 1, \quad Y := \frac{y}{2\varpi^3}.$$

Значения всех координат представлены в табл. 3.

Таблица 3

Точка	x	y	X	Y	z
$O = B^+$	∞	∞	∞	∞	0
B^-	-2	0	$\approx -6{,}875$	0	$\frac{\mathrm{i}}{2}$
A^-	-1	0	0	0	$\frac{1+\mathrm{i}}{2}$
A^+	0	0	$\approx 6{,}875$	0	$\frac{1}{2}$
O^+	$-2-\sqrt{2} \approx -3{,}414$	$(-2-\sqrt{2})\mathrm{i}$	$\approx -16{,}598$	$\approx -123{,}097\mathrm{i}$	$\frac{\mathrm{i}}{4}$
O^-	$-2-\sqrt{2} \approx -3{,}414$	$(2+\sqrt{2})\mathrm{i}$	$\approx -16{,}598$	$\approx 123{,}097\mathrm{i}$	$\frac{3\mathrm{i}}{4}$
$\underline{\infty}^+$	$-2+\sqrt{2} \approx -0{,}585$	$(2-\sqrt{2})\mathrm{i}$	$\approx 2{,}847$	$\approx 21{,}119\mathrm{i}$	$\frac{1}{2}+\frac{\mathrm{i}}{4}$
$\underline{\infty}^-$	$-2+\sqrt{2} \approx -0{,}585$	$-(2-\sqrt{2})\mathrm{i}$	$\approx 2{,}847$	$\approx -21{,}119\mathrm{i}$	$\frac{1}{2}+\frac{3\mathrm{i}}{4}$

В единичном квадрате на комплексной плоскости эти точки расположены так:



4.2. Факторизация $\mathbb{C} \to \frac{\mathbb{C}}{\Lambda}$

Поверим, что аналогичные преобразования определены и в общем случае, т. е. уравнения наших аффинных кубик $\dot{\mathbf{E}}$ имеют вид

$$Y^2 = 4X^3 - g_2X - g_3,$$

а проективные с уравнением

$$Y^2Z = 4X^3 - g_2XZ^2 - g_3Z^3$$

получаются заклеиванием единственного прокола $O = (0 : 1 : 0)$,

$$\dot{\mathbf{E}} = \mathbf{E} \setminus \{O\}.$$

Теорема. *Голоморфное отображение* $(\wp_\lambda, \wp'_\lambda) \colon \mathbb{C} \setminus \Lambda \to \dot{\mathbf{E}}$ *является неразветвлённым накрытием и продолжается до голоморфного отображения* $(\wp_\lambda : \wp'_\lambda : 1) \colon \mathbb{C} \to \mathbf{E}$. *Последнее отображение является гомоморфизмом групп* $(\mathbb{C}, +) \to (\mathbf{E}, \oplus)$ *с ядром* Λ.

Доказательство см. [4]. □

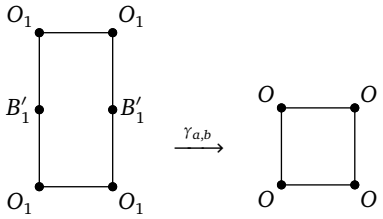
4.3. Подрешётки и изогении

Для $d \in \mathbb{N}$ назовём *d-изогенией* голоморфное отображение $\mathbf{E}' \to \mathbf{E}$ комплексных эллиптических кривых, которое является гомоморфизмом групп и при котором у каждой точки $P \in \mathbf{E}$ имеется ровно d прообразов.

Из приведённых результатов легко вытекает, что

если $\mathbf{E} = \frac{\mathbb{C}}{\Lambda}$ *— произвольная комплексная эллиптическая кривая, то d-изогении* $\mathbf{E}' \to \mathbf{E}$ *находятся в естественном взаимно однозначном соответствии с подрешётками* $\Lambda' \subset \Lambda$ *индекса* d.

При $d = 2$ с учётом *вещественной структуры* на $\mathbf{E}_{a,b}$ получаем обещанную визуализацию изогении Гаусса:



Удивительна её простота в сравнении с формулами, которые её определяют!

§ 5. Заключение

Приведя сначала прямое решение задачи о равенстве двух интегралов, оказавшееся необычайно громоздким, мы затем использовали эту задачу как повод рассказать о нескольких разделах классической математики в современных терминах, стараясь подтвердить слова Гаусса о *совершенно новой области анализа*, приведённые в конце § 1. Оказалось, что найденная Ланденом и Гауссом загадочная подстановка в эллиптическом интеграле может быть объяснена простыми и естественными конструкциями, если развить некоторые понятия топологии, алгебраической геометрии и теории функций комплексной переменной.

Автор надеется, что читатель, впервые встретившийся с этими понятиями, заинтересуется ими, но останется не удовлетворён уровнем изложения, принятым в статье, и захочет глубже изучить их. Многочисленные доступные в наше время учебники, статьи и материалы курсов разного уровня популярности предоставляют такую возможность.

Как и всякая хорошая задача, решённая в статье задача из «Математического просвещения» 2017 года, допускает разнообразные обобщения и аналоги, в том числе связанные с открытыми проблемами. Возможно, один из лучших способов освоить упомянутую в статье классику — включиться в активные исследования, продолжающиеся по сей день в *совершенно новой области анализа*.

Список литературы

[1] *Borwein J. M., Borwein P. B.* Pi and the AGM — A Study in Analytic Number Theory and Computational Complexity. New York: Wiley, 1987.

[2] *Cox D. A.* The arithmetic-geometric mean of Gauss // L'Enseignemant Mathématique. 1984. Vol. 30. P. 275–330.

[3] *Gauss C. F.* Werke, III. Göttingen, 1876.

[4] *Гурвиц А.*, *Курант Р.* Теория функций. М.: Наука, 1968.

[5] Математическое просвещение. Сер. 3. Вып. 21. М.: МЦНМО, 2017.

[6] *Острик В. В.*, *Цфасман М. А.* Алгебраическая геометрия и теория чисел: рациональные и эллиптические кривые. М.: МЦНМО, 2001.

[7] *Шафаревич И. Р.* Основы алгебраической геометрии. М.: МЦНМО, 2007.

Георгий Борисович Шабат, РГГУ
george.shabat@gmail.com