Лекция 5. Сложение точек на эллиптических кривых

1 Одна школьная задача

Задача (Задача 500 из сборника Шарыгина)**.** Существует ли неравнобедренный треугольник, такой, что точки пересечения биссектрисс его углов с противоположными сторонами образуют равнобедренный треугольник?

Оказывается, что такой треугольник на самом деле существует. Нетрудно понять, что если это так, то он будет тупоугольным. Явных же конструкций до некотрого времени придумано не было. Один из подходов к решению: записать соотношение между сторонами в виде уравнения и работать с ним. Такой подход приводит к ответу, однако никакой «наглядности» в таком подходе нет. Другое, совершенно неожиданное решение получается, если рассмотреть правильный 7-угольник. Однако, и в этом подходе ясно, что стороны треугольника иррациональны и являются громоздкими конструкциями из корней разных степеней (будь они квадратны, 7-угольник бы строился циркулем и линейкой). Вопрос, который можно поставить: а существует ли хоть один целочисленный треугольник с таким свойством.

Компьютерный перебор показывает, что среди целых чисел до 10^6 искомой комбинации нет. С другой стороны, если посмотреть на условие задачи, совершенно ясно, что указанное свойство инвариантно относительно гомотетии, а значит, задача состоит в поиске рациональных точек на какой-то плоской кривой, задаваемой каким-то алгебраическим уравнением $\psi(a,b)=0$. Может показаться странным, что на всей этой кривой нет рациональных точек и эти сомнения оказываются небеспочвенны.

Вычислив отрезки, на которые разбивают биссектриссы каждую сторону и записав теорему косинусов, можно получить следующее соотношение на стороны треугольника:

$$\cos\varphi=\frac{a^2+c^2-b^2}{2ac},\cos\psi=\frac{b^2+c^2-a^2}{2bc}$$

$$\left(\frac{ac}{b+c}\right)^2+\left(\frac{ac}{a+b}\right)^2-2\frac{ac}{b+c}\frac{ac}{a+b}\frac{a^2+c^2-b^2}{2ac}=$$
$$\left(\frac{bc}{a+c}\right)^2+\left(\frac{bc}{a+b}\right)^2-2\frac{bc}{a+c}\frac{bc}{a+b}\frac{b^2+c^2-a^2}{2bc}.$$

Упражнение 1. Упростить соотношение до вида:

$$\frac{a}{b+c}+\frac{b}{a+c}=\frac{c}{a+b}$$

У этого соотношения безусловно есть «лишние решения», соответствующие разным вырожденным случаям, равнобедренным треугольникам, а также тем a,b,c, которые вообще не задают никакой треугольник.

Упражнение 2. Получить другие формы того же соотношения:

- $(a+b+c)(a^2+b^2+c^2)+abc=0$
- $\frac{a(c-a)}{(b+c)^2}=\frac{b(c-b)}{(a+c)^2}$
- $c^3+c^2(a+b)=c(a^2+ab+b^2)+(a^3+a^2b+ab^2+b^3)$

Еще одна параметризация кривой, хорошая тем, что в ней можно не беспокоиться о соблюдении неравенства треугольника, связана с отрезками, на которые разбиваются стороны высотами, проведёнными из точки пересечения биссектрисс. Пусть эти отрезки отрезки равны x, y, z, тогда соотношение переписывается с учётом того, что $a=x+y, b=x+z, c=y+z$:

$$4x^3+9x^2y+9x^2z+6xyz+5xy^2+5xz^2-3y^2z-3yz^2=0.$$

1.1 Проективная плоскость и проективизация пространства

При изучении однородных уравнений произвольной степени естественным образом получается, что вместе с любым решением появляются все, отличающиеся от него в константное число раз. Путём отождествления всех решений, отличающихся в константное число раз, происходит *проективизация* пространства решений. Таким образом кривые, задаваемые однородными формами, образуют проективное пространство, так как получаются путем проективизации линейного пространства всех форм.

Однако проективные пространства, хоть и представляют собой более точную модель, обладают меньшей геометрической наглядностью. Поэтому часто рассматривают более простую вещь — отождествление лишь тех решений, которые лежат на одном луче, проходящем через точку 0. Полученный объект гораздо более интуитивен, так как представляет собой просто точки на сфере, являющиеся решением данного уравнения. В этом случае, конечно, всегда нужно помнить, что вместе с любым решением входит и диаметрально ему противоположное.

Непосредственно на сфере или в проективной записывать уравнения не слишком удобно. Поэтому используется метод афинных карт. В общем случае, выбирается афинное подпространство (гиперплоскость) размерности на 1 меньше, чем все пространство, не проходящая через 0, и каждой точке сопоставляется её центральная проекция на выбранное подпространство. Естественно, что точки, лежащие в плоскости, параллельной данной и проходящей через 0 не могут быть центрально спроектированы (проектируются на бесконечно удалённую прямую) и не представлены координатно на выбранной карте.

Однако, то, что некоторые решения выпадают из рассмотрения на самом деле не представляет собой большой проблемы. Однородное уравнение, получаемое в случае, если координаты решения лежат в параллельной карте плоскости (для простоты можно представить, что рассматривается

карта $z = 1$, что соотвествует делению уравнения на z в соответствующей степени и отдельному рассмотрению случая $z = 0$) имеет не более, чем конечное (ограниченное степенью формы) множество решений.

На этом наблюдении основан один трюк, который оказался довольно успешным при исследовании рациональных точек на произвольных кривых, заданных алгебраическими уравнениями. Стандартное исследование симметрий уравнения $\varphi(x, y) = 0$ подразумевает, например, исследование афинных преобразований, упрощающих вид уравнения. Однако, если добавить дополнительную переменную z и дополнить умеющуюся форму до однородной, домножив каждое слагаемое на z в нужной степени, то множество простых преобразований, с помощью которых можно упрощать уравнение сильно обогащается: в частности очень полезным является домножение на какую-либо матрицу 3×3. С другой стороны, из предыдущих рассуждений получается, что при таких преобразованиях «потеряно» может быть лишь конечное число решений. Сила проективных преобразовний состоит в том, что многие случаи, возникающие при рассмотрении других групп (евклидовых или афинных преобразований) перестают быть отдельными случаями в проективном случае.

В качестве примера можно привести поиск рациональных точек на кривой, заданной квадратичной формой $Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0$. Принципиально различны в проективном подходе следующие три случая: форма задаёт две совпадающие прямые, две пересекающиеся прямые или форма не раскладывается в произведение линейных сомножителей и задаёт эллипс. Случаи параллельных прямых и разных видов коник оказываются автоматически разобранными, эквивалентными уже перечисленным (так как любые две прямых на проективной плоскости пересекаются, а вид коники определяется просто выбором карты).

Строгое утверждение состоит в том, что любое однородное уравнение $Ax^2 + Bxy + Cy^2 + Dxz + Eyz + Fz^2 = 0$, которое не раскладывается в произведение двух линейных форм, может быть переведено преобразованием

$$\begin{pmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

в уравнение вида $\widetilde{x}\widetilde{z} = \widetilde{y}^2$, которое обладает рациональной параметризацией, то есть отображение $\mathbb{PQ}^2 \to \mathbb{PQ}^3$, заданное формулой $\widetilde{x} = u^2, \widetilde{y} = uv, \widetilde{z} = v^2$, полностью описывает рациональные решения уравнения кроме, быть может, двух.

1.2 Получение одной рациональной точки на кривой

Возвращаясь к уравнению конкретной кривой, записанной в координатах x, y, z, можно отметить примечательную его особенность: будучи записанной в карте y, z, то есть в координатах y, z при $x = 1$, оно принимает вид $4 + 9y + 9z + 6yz + 5y^2 + 5z^2 - 3y^2z - 3yz^2 = 0$. Можно заметить, что по каждой координате в отдельности оно является квадратным.

Тогда применим следующий метод: возьмём какую-либо нетривиальную целую точку, например, подойдет $(y, z) = (-3, 1)$. При $y = -3$ уравнение обращается в квадратное, притом известен его рациональный коренью. Это значит, что найденный второй корень также будет рациональным. Перейдя ко второму корню, можно получить третий, зафиксировав z и найдя соотвествующее y. Итерируя этот процесс, можно получать новые рациональные точки на кривой до тех пор, пока не будет обнаружен цикл.

Таким способом на четвертой итерации получается точка (y_4, z_4), причём обе её координаты положительны. Это значит, что она даёт решение исходной задачи. Если выразить его в исходный координатах, получится:

$$a = 1481089,\ b = 18800081,\ c = 19214131.$$

Однако такой метод анализа больше похож на некоторое ad-hoc наблюдение, которое к тому же не позволяет получить характеризацию всех точек на этой кривой, поэтому важно уделить внимание тому, как тот же результат ложится в рамки более общего подхода к анализу эллиптических кривых.

2 Эллиптические кривые

2.1 Понятие эллиптической кривой, особые точки и точки перегиба

При исследовании кривой, заданной однородной 3-формой $\Phi(a, b, c) = 0$ выделяется несколько случаев вырожденного поведения. Первый из них заключается в том, что форма Φ факторизуется, то есть представляется в виде произведения двух форм меньшей степени. Этот случай менее интересен, так как кривая, задаваемая такой формой распадается в объединение прямой и коники или даже просто в три прямые.

Второй, более тонкий вид вырождения можно рассмотреть на примере кривой, задаваемой уравнением $y^2 = x^3$ (проективной формой $\Phi(a, b, c) = ac^2 - b^3$). На аналитическом языке можно сказать, что кривая имеет излом в точке $(0, 0)$, однако в алгебраической геометрии принято формулировать это иначе (это связано с желанием работать в более общем случае, нежели просто над полем $\mathbb{R}$ или $\mathbb{C}$). Можно заметить, что любая прямая, проходящая через $(0, 0, 0)$, является касательной к кривой, то есть $\Phi(a + \alpha t, b + \beta t, c + \gamma t)$ делится на t^2 для любых (α, β, γ). Точка иррегулярности такого рода называется *особой*, а кривые, содержащие особые точки носят название *особых* кривых.

Замечание. Поиск особых точек на кривой эквивалентен решению системы уравнений $\frac{\partial \Phi}{\partial a} = \frac{\partial \Phi}{\partial b} = \frac{\partial \Phi}{\partial c} = 0$.

Упражнение 3.

- Показать, что форма кривой из задачи про биссектральный треугольник неразложима.

- Показать, что эта кривая не является особой.

Определение 1 (Основное определение теории элиптических кривых)**.** Кривая $\Phi(a, b, c) = 0$ третьей степени называется *эллиптической* над полем K называется эллиптической, если она неразложима, неособа и существует хотя бы одна точка, все координаты которой лежат в K.

Замечание. Исследуемая кривая содержит пару рациональных точек $(1, 0, 1)$ и $(0, 1, 1)$, а значит, по упражнению 3 является эллиптической.

Основным инструментом получения точек на эллиптических кривых является следующая операция, называющаяся *сложением* точек: имея две точки на кривой, можно провести через них прямую. Точка её пересечения с эллиптической кривой* будет рациональна, если исходные две точки были таковыми.

В частности, между известными точками $(1, 0, 1)$ и $(0, 1, 1)$ должна находиться одна несовпадающая с ними точка (так как уравнение кривой инвариантно при замене a и b). Эта точка $(1, -1, 0)$. Стоит отметить, что эта точка инвариантна относительно замены a и b, так как переходит в точку, получающуюся домножением на -1, то есть эквивалентна ей в проективном смысле.

Вторым способом получения рациональных точек на кривой является проведение касательной в какой-то известной рациональной точке.

Замечание. Точки перегиба кривой задаются уравнением

$$\det \begin{pmatrix} \frac{\partial^2 \Phi}{\partial a^2} & \frac{\partial^2 \Phi}{\partial a \partial b} & \frac{\partial^2 \Phi}{\partial a \partial c} \\ \frac{\partial^2 \Phi}{\partial b \partial a} & \frac{\partial^2 \Phi}{\partial b^2} & \frac{\partial^2 \Phi}{\partial b \partial c} \\ \frac{\partial^2 \Phi}{\partial c \partial a} & \frac{\partial^2 \Phi}{\partial c \partial b} & \frac{\partial^2 \Phi}{\partial c^2} \end{pmatrix} = 0.$$

Указанная матрица называется *гессианом* формы (для 3-формы её определитель также является 3-формой). В точке, где $\det H = 0$, можно найти также направление перегиба v, такое, что $Hv = 0$.

Упражнение 4. $(1, -1, 0)$ — точка перегиба для данной кривой с направлением перегиба $(1, 1, -4)$.

Если на кривой найдена точка перегиба, то бывает разумно рассмотреть связанный с ней базис: выбрать в качестве первого вектора радиус-вектор точки, направление перегиба взять за второй элемент базиса, а третий вектор достроить перпендикулярно первым двум. В этом базисе точка перегиба перейдет в бесконечно удалённую точку, а вектор перегиба будет сонаправлен с бесконечно удалённой прямой.

*Для эллиптической кривой такая точка будет единственна в силу простого варианта теоремы Гильберта о нулях: если форма степени три принимает значение 0 хотя бы в четырёх точках данной прямой, то она принимает нулевое значение в каждой точке прямой и её форма делится на уравнение прямой в алгебре многочленов, что исключено, так как формы эллиптических кривых неразложимы

Упражнение 5. Переписать уравнение кривой в базисе для точки перегиба $(1, -1, 0)$.

Это упражнение на самом деле является частным случаем гораздо более общего факта. А именно, любую эллиптическую кривую можно привести к нормальной форме Вейерштрасса:

$$y^2 + axy + by = x^3 + \alpha x^2 + \beta x + \gamma.$$

В случае, если действие происходит в поле $\mathbb{Q}$, этот вид можно упростить ещё больше, получив выражение:

$$y^2 = x^3 + Ax + B,$$

притом A и B определяются однозначно.

2.2 Свойства операции сложения

Пусть операция сложения двух точек A, B на эллиптической кривой обозначена как $A + B$, а точка перегиба обозначена нулём*. Тогда несложно проверить, что $0 + 0 = 0$ и $A + 0 = \widetilde{A}$, где $\widetilde{A}$ — точка, симметричная к A относительно точки перегиба. Следуя естественному желанию превратить точку 0 действительно в 0 какой-то абелевой группы относитльно операции сложния, разумно переопределить сумму двух точек следующим образом:

Определение 2. *Суммой* точек A и B на эллиптической кривых называется точка C, полученная следующим построением: построить третью точку D на прямой AB, и взять в качестве C третью точку на прямой $0D$, где 0 — точка перегиба.

Замечание. При таком определении можно в самом деле убедиться, что $0 + 0 = 0, A + 0 = A$, а также, что $\exists \overline{A} : A + \overline{A} = 0$ и $A + B = B + A$. Для группы осталось только проверить ассоциативность сложения.

Две великие по силе и красоте теоремы, полученные около 100 лет назад, дают практически исчерпывающую характеризацию рациональных точек на эллиптических кривых.

Теорема 1 (Пуанкаре)**.** *Сложение точек на любой эллиптической кривой ассоциативно.*

Теорема 2 (Морделла)**.** *Абелева группа рациональных точек на любой эллиптической кривой конечнопорождена.*

*Замечание, которое нужно сделать в общем случае для анализа произвольной эллиптической кривой: точки перегиба кривой не обязаны быть рациональными (вещественная точка перегиба у кривой третьего порядка есть всегда). В этом случае операцию сложения, определенную ниже тем не менее можно определить относительно произвольной рациональной точки на кривой, однако, самым простым для анализа случаем является наличие рациональной точки перегиба. Общие результаты, описанные ниже, тем не менее, выполнены и в общем случае

2.3 Примерный анализ группы исследуемой кривой и гипотезы

Пользуясь этими мощными инструментами, можно исследовать данную в задаче эллиптическую кривую. Так, на ней можно найти элемент кручения D : $D+D=0$ (касательная, проведённая в такой точке должна пересекать кривую в точке её перегиба) и порождающий элемент бесконечного порядка, что даёт право предположить, что группа рациональных точек кривой изоморфна $\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})$.

Найденные образующие: $A = (1, 0, -1), D = (1, 1, -1)$, порядок первой бесконечен, вторая образует кручение степени 2. Параллель с поиском рациональной точки в других координатах путём отражения с фиксированной одной координатой по-видимому состоит в том, что одна результат n-й итерации есть $(2n + 1)A + D$, то есть точка, полученная на 4-й итерации того алгоритма должна быть равна $9A + D$, что верно и проверяется непосредственно.

Другие наблюдения о данной кривой, которые можно развивать: число $6A$ очень близко к D. Если к тому же $(12n + 6)A \to D$ при $n \to \infty$, то это может стать отправной точкой для доказательства бесконечности числа целочисленных треугольников, подходящих под условие задачи.

3 Доказательство теоремы Пуанкаре

Пусть даны точки A, B, E на эллиптической кривой. Пусть также

- R — третья точка на прямой A, B, то есть $A + B + R = 0$
- L_1 — прямая, содержащая A, B, R (обозначение $L_1 = ABR$)
- $L_2 = R0\overline{R}$, где $R + \overline{R} = 0$ (ясно, что $\overline{R} = A + B$)
- $L_3 = E\overline{R}S$
- $L_4 = S0\overline{S}$, $\overline{S} = (A + B) + E$
- $M_1 = BEQ$
- $M_2 = Q0\overline{Q}$
- $M_3 = A\overline{Q}T$
- $M_4 = T0\overline{T}, \overline{T} = A + (B + E)$

Таким образом, необходимо показать, что $S = T$ или, что тоже самое $\overline{S} = \overline{T}$ (то есть в доказательстве использовать прямые L_4 и M_4 не придётся). Итого, получилось 10 точек на кривой: $A, B, E, O, R, \overline{R}, Q, \overline{Q}, S, T$.

Упражнение 6. Если $A, B, E, 0$ — различны, то никакие две точки из наборов $\{A, B, \ldots, \overline{Q}\} \cup \{S\}$, $\{A, B, \ldots, \overline{Q}\} \cup \{T\}$ не совпадают.

Утверждение 1. Пусть форма $\Phi(x, y, z)$ степени d задаёт кривую $\Phi(x, y, z) = 0$. Пусть также нашлась такая коника, на которой можно выбрать $2d + 1$ точек, лежащих на кривой. Тогда $\Phi(x, y, z)$ делится на уравнение коники как многочлен.

Доказательство. Сначала можно показать, что $\Phi(x, y, z) = 0$ на любой точке коники. В самом деле, как было выяснено, в проективных координатах коника приводится к виду $xz = y$ и параметризуется с помощью u, v вектором (u^2, uv, v^2). Тогда, при подстановке в уравнение кривой получается, что однородный многочлен степени $2d$ равен нулю в $2d + 1$ точке, откуда следует, что он есть тождественный ноль.

Тогда многочлен Φ представляется в виде $\Phi_1 \cdot (zx - y^2) + \Phi_2(x, z) + y\Phi_3(x, z)$ и в силу наложенных ограничений $\Phi_2 \equiv \Phi_3 \equiv 0$ (второе слагаемое содержит только чётные степени u, v, третье только нечётные). □

Замечание. Совершенно аналогично, если прямая, заданная линейной формой $L = 0$, содержит $d + 1$ точку, на которой $\Phi(x, y, z) = 0$, то Φ делится на L как многочлен.

Пусть теперь даны k точек $P_1, \ldots, P_k$ и C^2_{d+2}-мерное пространство всех форм степени d. Тогда подпространство, задаваемое линейными уравнениями на коэффициенты формы $\Phi(P_1) = 0, \ldots, \Phi(P_k) = 0$, имеет размерность $\dim V \geqslant C^2_{d+2} - k$.

Утверждение 2. Если точки $P_1, \ldots, P_5$ таковы, что никакие 4 из них не лежат на одной прямой, то пространство всех 2-форм, зануляющихся на этих точках одномерно.

Доказательство. Первый случай: если никакие три точки не лежат на одной прямой. Тогда если $\dim V > 1$, то существовали бы две различные коники, обнуляющиеся на этих точках. Тогда так как при $d = 2$ выполнено $2d + 1 = 5$, то их уравнения делятся друг на друга в силу утверждения 1, что ведёт к противоречию. □

Упражнение 7. Разобрать случай, когда три точки лежат на одной прямой.

Утверждение 3. Если точки $P_1, \ldots, P_8$ таковы, что никакие 7 из них не лежат на невырожденной конике и никакие 4 из них не лежат на одной прямой. Тогда размерность пространства 3-форм, проходящих через эти точки равна 2.

Доказательство. Первый случай: пусть никакие 6 не лежат на невырожденной конике и никакие 3 не лежат на одной прямой. Тогда через первые 5 точек проведем конику (она будет невырождена, так как никакие 3 не лежат на одной прямой). Тогда если размерность пространства форм больше 2, то при наложении ещё два линейных ограничений $\Phi(Q) = 0$, для каких-то двух точек Q_1, Q_2 на конике получится пространство размерности не меньше 1, то есть существует кривая степени 3, проходящая через $P_1, \ldots, P_8, Q_1, Q_2$.

Тогда так как $2 \cdot 3 + 1 = 7$, то $\Phi = Q \cdot L$, где Q — уравнение коники. Но тогда L принимает нулевое значение на оставшихся 3 из 10 точек, то есть они лежат на одной прямой, противоречие.

Пусть теперь существуют 6 точек, лежащих на невырожденной конике. Тогда если к ним добавить еще одну и доказать, что полученное пространство одномерно, то из этого будет следовать утверждение. Тогда существует кривая второго порядка, проходящая через все эти 9 точек, в том числе через 7 точек коники, значит $\Phi = QL$, где Q —уравнение коники. Оставшаяся часть L проходит через две остальные точки, то есть задана однозначно, значит существует единственная с точностью до пропорциональности форма, проходящая через эти 8 точек, что и доказывает одномерность этого пространства, то есть исходное двумерно.

Последний случай: есть 3 точки на одной прямой. Тогда можно дополнить их четвертой точкой и форма будет делиться на уравнение полученной прямой L, то есть $\Phi = QL$, причём форма Q содержит оставшиеся 5 точек, то есть строится однозначно. Аналогично предыдущему пункту, пространство 2-форм, проходящих через исходные 8 точек двумерно. □

Пусть теперь C — исходная эллиптическая кривая, $D_1 = L_1 M_2 L_3$ (как произведение форм), $D_2 = M_1 L_2 M_3$. Несложно заметить, что на D_1 лежат 9 из точек $A, B, E, O, R, \overline{R}, Q, \overline{Q}, S, T$, притом в их числе точка S, на D_2 аналогично лежат 9 точек, в том числе T. Итак, C и D_1 имеют 9 общих точек и, поскольку C невырожденна, больше точек пересечения нет и более того, никакие 7 не лежат на одной конике и никакие 4 не лежат на одной прямой. Значит по утверждению, C, D_1 образуют базис в пространстве 3-форм, проходящих через первые 8 точек. Так как D_2 тоже проходит через эти 8 точек, то $D_2 = \lambda C + \mu D_1$. Но точка S лежит на C и на D_1, значит и на D_2. Итак, 10 точек лежат на трёх прямых M_1, L_2, M_3 и значит $S = T$. Доказательство завершено.