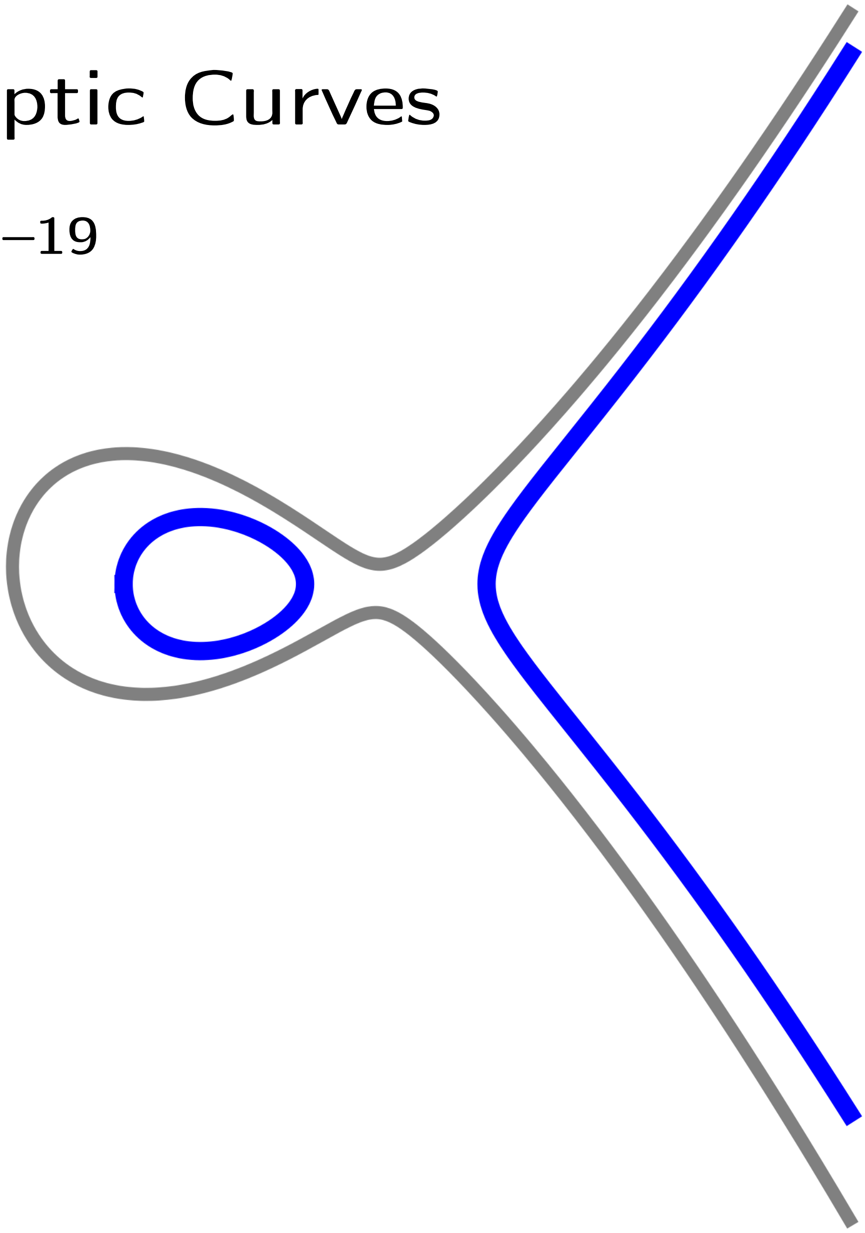
G13ELL=MATH3031

Elliptic Curves

2018–19



Chris Wuthrich

Essential information

Lecturer : Chris Wuthrich,
christian.wuthrich@nottingham.ac.uk,
phone 14920.

Lectures :
- Tuesdays 10 am – noon, Physics C29
- Thursdays noon – 1pm, Pharmacy A6
- Fridays 11 am – noon, Maths A17

Office Hours : On Thursday mornings in my office C58 in the Maths building. If you wish to meet me at any other time, just try knocking at the door or – if you want to make certain that I am in – please contact me by email or by phone.

Module webpage : The module webpage is on moodle:

`http://moodle.nottingham.ac.uk/course/view.php?id=68482.`

Lecture notes : On the moodle page you will find all these lecture notes including all the pictures. The printed version omits pictures that you are asked to draw during lectures.

Booklist : There are plenty of books and online lecture material on elliptic curves. This module recommends [6], [1] and [7] (in the list on page 4) as the best books to consult. Please ask if you are interested in a particular topic.

Problem classes : We will have problem classes, in average one per week. During this hour you will work (with my help) on exercises relating to the lectures. There are also unassessed coursework sheets which you are asked to hand in your solutions for marking.

Assessment : There is a 3-hour exam in January. There will be FIVE questions in the exam paper. Credit will be given for the best FOUR answers. Consult the moodle page for the exam paper of last year. More information will be given later.

Computer software : Not needed at all, but if you wish to experiment with elliptic curves you may want to try out the free SageMath. You may use it freely on CoCalc. All graphics and computations in these notes are done with SageMath.

Contents

References

[1] J. W. S. Cassels, *Lectures on elliptic curves*, London Mathematical Society Student Texts, vol. 24, Cambridge University Press, Cambridge, 1991, George Green Library, QA565 CAS.

[2] H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel, and R. Remmert, *Numbers*, Graduate Texts in Mathematics, vol. 123, Springer-Verlag, New York, 1991.

[3] David Eisenbud, Mark Green, and Joe Harris, *Cayley-Bacharach theorems and conjectures*, Bull. Amer. Math. Soc. (N.S.) **33** (1996), no. 3, 295–324.

[4] Kenneth Ireland and Michael Rosen, *A classical introduction to modern number theory*, second ed., Graduate Texts in Mathematics, vol. 84, Springer-Verlag, New York, 1990.

[5] Franz Lemmermayer, *Elliptic curves, historical remarks*, Avaialble at `http://www.fen.bilkent.edu.tr/~franz/ta/ta01.pdf`, last visited Sep 2018.

[6] Joseph H. Silverman and John T. Tate, *Rational points on elliptic curves*, second ed., Undergraduate Texts in Mathematics, Springer, Cham, 2015, George Green Library, QA565 SIL.

[7] Lawrence C. Washington, *Elliptic curves*, second ed., Discrete Mathematics and its Applications (Boca Raton), Chapman & Hall/CRC, Boca Raton, FL, 2008, George Green Library, QA565 WAS and QA567.2.E44 WAS.

1 Chords and tangents

We start this module on elliptic curves with some examples. The motto here is “Geometry helps solving equations”.

Say, we want to solve the equation

$$C: \qquad x^2 + y^2 - 2x = 0; \tag{1.1}$$

however if possible without using square roots or non-algebraic functions like sin and log. We start by making a drawing as in Figure 1. Since the equation can be reformulated as $(x-1)^2 + y^2 = 1$ we see a circle of radius 1 centred at $(1, 0)$. For instance $P = (0, 0)$ is an obvious solution. Pass a line through P. It can be given by the equation

$$L: \qquad y = mx$$

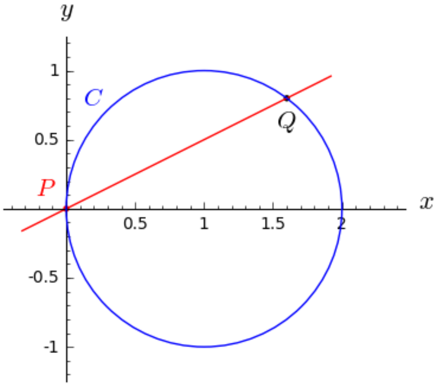


Figure 1: A geometric picture of our equation (1.1)

where m is some constant. The points in the intersection $L \cap C$ satisfy

$$\begin{aligned} x^2 + (mx)^2 - 2x = 0 \quad &\Rightarrow \quad \bigl(1+m^2\bigr)\, x^2 - 2x = 0 \\ &\Rightarrow \quad x \cdot \bigl(\bigl(1+m^2\bigr)\, x - 2\bigr) = 0 \end{aligned}$$

If $x = 0$, then $y = m \cdot 0 = 0$ and we get back the point P. Of course this is no surprise as P was on C and L. Otherwise, if $x \neq 0$, we have

$$Q = (x, y) = \Bigl(\frac{2}{1+m^2}, \frac{2m}{1+m^2}\Bigr). \tag{1.2}$$

Is Q really on C? Well, ...

$$\Bigl(\frac{2}{1+m^2}\Bigr)^2 + \Bigl(\frac{2m}{1+m^2}\Bigr)^2 - 2\frac{2}{1+m^2} = \frac{2 + 4m^2 - 2(1+m^2)}{(1+m^2)^2} = 0 \tag{1.3}$$

so yes, Q is in $C \cap L$. Therefore we found plenty of solutions to C, but did we find all of them? Well, if (x_0, y_0) is a solution to C with $x_0 \neq 0$, then there is an $m = y_0/x_0$ such that the line $y = mx$ passes through P and (x_0, y_0). Therefore we will recover any solution to C other than P from the parametrisation (1.2). We can write

$$C(\mathbb{R}) = \Bigl\{\Bigl(\frac{2}{1+m^2}, \frac{2m}{1+m^2}\Bigr) \Bigm| m \in \mathbb{R}\Bigr\} \cup \bigl\{(0,0)\bigr\}$$

where we have denoted by $C(\mathbb{R})$ the solution set for C with coordinates x and y in $\mathbb{R}$.

The real numbers? Yes, when we relied on our picture we really made a picture for the real solutions. But what if we would like to find all solutions to (1.1) with coordinates in $k = \mathbb{Q}$ or $k = \mathbb{C}$. Does the above still work? First of all the computation (1.3) works for any field k and $m \in k$ as long as $m^2 \neq -1$. So no matter what k is, we find a parametrisation. For any solution (x_0, y_0) with $0 \neq x_0 \in k$ and $y_0 \in k$ we can set $m = y_0/x_0$ and thus we recover all solutions except when $m^2 = -1$. For instance we get that

$$C(\mathbb{Q}) = \Bigl\{ \Bigl(\frac{2}{1+m^2}, \frac{2m}{1+m^2} \Bigr) \Bigm| m \in \mathbb{Q} \Bigr\} \cup \Bigl\{ (0,0) \Bigr\}.$$

It even works for $k = \mathbb{F}_3 = \mathbb{Z}/3\mathbb{Z}$, the field of three elements:

$$\begin{aligned} C(\mathbb{F}_3) &= \Bigl\{ \Bigl(\frac{2}{1+m^2}, \frac{2m}{1+m^2} \Bigr) \Bigm| m \in \mathbb{F}_3 \Bigr\} \cup \Bigl\{ (0,0) \Bigr\} \\ &= \Bigl\{ (2,0), (1,1), (1,2), (0,0) \Bigr\}. \end{aligned}$$

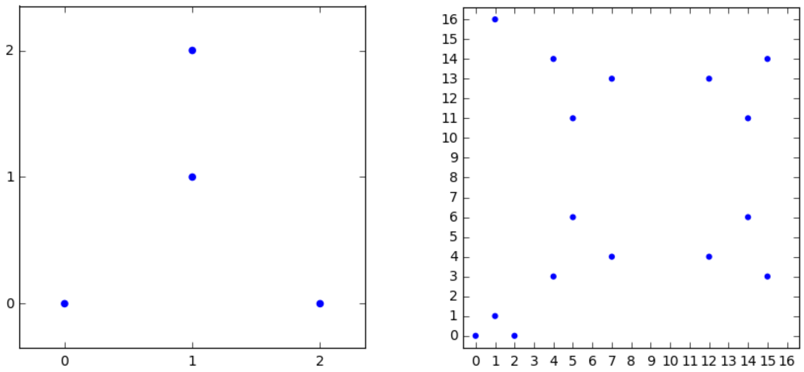


Figure 2: The "circle" (1.1) over $\mathbb{F}_3$ and over $\mathbb{F}_{17}$

However this time the picture is quite far from a circle. See Figure 2. However the geometric method helped us to find all solutions quicker than if we checked all possible x and y in $\mathbb{F}_3$.

We pass to a more complicated example. In the above the equation was of degree 2 and we will later see that the method used there generalises. Our next example is of degree 3, it will turn out to be an example of an elliptic curve. The equation is

$$C : \qquad x^3 + y^3 = 1729 \tag{1.4}$$

and we are looking for solutions with coordinates in $\mathbb{Q}$. One may be able to spot[*] some solutions, including $P=(1,12)$ and $Q=(9,10)$. See Figure 3. Let

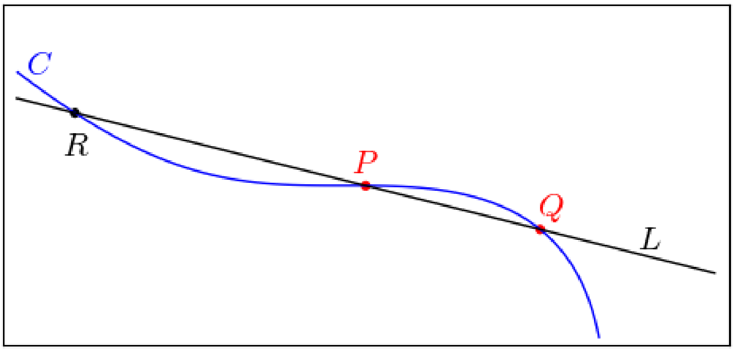


Figure 3: Chords on the taxicab curve

us pick a general line through P

$$L: \qquad y-12=m(x-1)$$

for some $m\in\mathbb{Q}$. The points in the intersection $C\cap L$ satisfy

$$\begin{aligned} 0 &= x^3+(mx-m+12)^3-1729 \\ &= (1+m^3)x^3+3m^2(12-m)x^2+3m(m^2-24m+144)x \\ &\qquad -m^3+36m^2-432m-1 \\ &= (x-1)\bigl((1+m^2)x^2-(2m^3-36m^2-1)x+m^3-36m^2+432m+1\bigr) \end{aligned}$$

where, in the last line, we found a factorisation because we knew that $x=1$ must be a solution, given that $P\in C\cap L$. Now we are stuck as we are left with a quadratic polynomial that does not seem to factor.

All is not lost. Let L be the line through P and Q. This is a particular line for which we know a second solution. The line corresponds to $m=-\frac{1}{4}$ above. We find

$$0=x^3+\bigl(-\tfrac{x-1}{4}+12\bigr)^3-1729=\tfrac{21}{64}(x-1)(x-9)(3x+37).$$

[*] 1729 is a famous taxicab number. Hardy visited Ramanujan in 1919. He wrote: "I remember once going to see him [Ramanujan] when he was lying ill at Putney. I had ridden in taxi-cab No. 1729, and remarked that the number seemed to be rather a dull one, and that I hoped it was not an unfavourable omen. 'No', he replied, 'it is a very interesting number; it is the smallest number expressible as the sum of two [positive] cubes in two different ways.' " Source: Wikipedia

This time, we find a third solution $x = -\frac{37}{3}$ and from the equation of the line we get $y = \frac{46}{3}$. This is indeed a point $R = (\frac{37}{3}, \frac{46}{3})$ in $C \cap L$ with rational coordinates. This method is called a "chord", it is a line from one point on the curve to another point on the curve. See Figure 3 for an illustration.

There is a second method to get a new solution, the "tangent" method. We can take the tangent T at P to the curve C; see Figure 4 for an illustration. We present here what would happen with this computation; though they are a bit tedious to do without the help of a computer. One can parametrise the tangent by

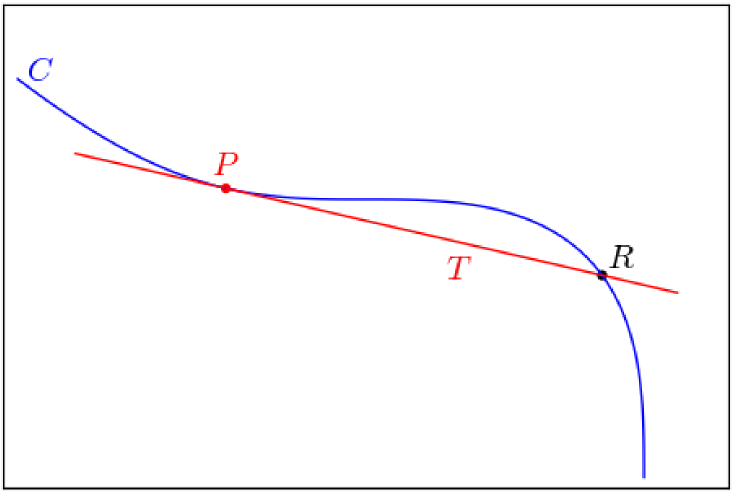


Figure 4: Using tangents to find new points

$$\ell(t) = \begin{pmatrix} 1 \\ 12 \end{pmatrix} + t \begin{pmatrix} 144 \\ -1 \end{pmatrix}$$

and the points in the intersection correspond to t such that

$$0 = (1 + 144t)^3 + (12 - t)^3 - 1729 = 2985983 \cdot t^2 \cdot \bigl(t + \tfrac{36}{1727}\bigr).$$

It is good that we find a factor t^2 here: The tangent at P has a sort of a double intersection with C. At this stage "double intersection" is a vague idea that we will turn into a rigorous notion later. We find the new point $\bigl(-\frac{3457}{1727}, \frac{20760}{1727}\bigr)$ on C with coordinates in $\mathbb{Q}$. Now we can apply the same constructions on and on again. In general, if $(x, y) \in C(\mathbb{Q})$, then the tangent procedure shows that the point

$$\Bigl(\frac{x(x^3 + 2y^3)}{x^3 - y^3}, \frac{y(2x^3 + y^3)}{y^3 - x^3}\Bigr)$$

is also in $C(\mathbb{Q})$. This allows us in a few steps to produce new solutions like

$$\Bigl(\frac{8062154141168529647191916 81}{6717887337780573931276140 9}, \frac{5526295949142309247137748 0}{6717887337780573931276140 9}\Bigr).$$

Visibly there seem to be infinitely many solutions now to our equation. However this is not obvious because it could happen that, after a while, both methods of chords and tangents start to reproduce only points that we found already. Through numerical experiment it seems however that the size of the numerator and denominator of our solutions grow very quickly so it seems impossible that there are finitely many solutions.

The main aim of this module is to investigate this chord and tangent idea on cubic curves. We will at some point specialise to certain equations of the form

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

where A and B are two integers. They will be called elliptic curves in Weierstrass form and they carry a very rich structure. In fact, we will use the above chord and tangent method to define an abelian group law on the set of solutions.

Let us end this initial section by explaining that cubic equations take a special place. If we had a general equation of degree four or higher then we cannot find any more a procedure using chords or tangents to produce new solutions from old ones. There are some exceptions to this, but the vast majority of higher degree curves are beyond the scope of what we can do here. A very surprising and extremely difficult result was shown by Gerd Faltings in the 1980ies: The solution set $C(\mathbb{Q})$ for such curves of higher degree[†] is always finite.

[†] Technically it is not the degree but the so-called genus that has to be large, larger than 1 in fact. Some higher degree curves have a small genus, but most have very large genus. The most general definition of elliptic curves requires the curve to have genus 1.

2 Groups and fields

Visibly there will be a question of fields involved. Later we will have a group law on a curve. We start with recalling the basic definition of these from G12ALN=MATH2015.

2.1 Abelian groups

Definition. An **abelian group** $\langle A, +\rangle$ is a non-empty set A together with an operation $+ : A \times A \to A$ satisfying the following axioms:

(G1 Closure): For all a, $b \in A$, we have $a + b \in A$.

(G2 Associativity): $(a + b) + c = a + (b + c)$ for all a, b, $c \in A$.

(G3 Identity): There is an element $0 = 0_A$ such that $a + 0 = 0 + a = a$ for all $a \in A$.

(G4 Inverses): For each $a \in A$ there is a unique element $-a \in A$ such that $a + (-a) = (-a) + a = 0$.

(G5 Abelian): For all a, b in A, we have $a + b = b + a$.

This is copied from Section 1.1 in G12ALN=MATH2015 (with the exception that groups were allowed to be non-abelian there, we will only have abelian ones here). The main examples used in this modules are $\langle \mathbb{Z}, +\rangle$ and the cyclic group $\langle {}^{\mathbb{Z}}/_{m\mathbb{Z}}, +\rangle$ of order m.

What do we need to know about abelian groups?

Theorem 2.1. *Let A be a finite abelian group. Then there are integers $n_1 > 0$, $n_1 > 0$, ..., $n_t > 0$ and prime numbers (not necessarily distinct) p_1, p_2, ..., p_t, such that*

$$A \cong {}^{\mathbb{Z}}/_{p_1^{n_1}\mathbb{Z}} \times {}^{\mathbb{Z}}/_{p_2^{n_2}\mathbb{Z}} \times \cdots \times {}^{\mathbb{Z}}/_{p_t^{n_t}\mathbb{Z}}$$

Up to reordering the terms the factorisation into cyclic groups is unique. One can use the Chinese remainder theorem to regroup them if necessary: If n and m are coprime integers then ${}^{\mathbb{Z}}/_{nm\mathbb{Z}} \cong {}^{\mathbb{Z}}/_{n\mathbb{Z}} \times {}^{\mathbb{Z}}/_{m\mathbb{Z}}$.

Given an abelian group of finite order, one can often detect the above factorisation if one knows who many elements of a particular order there are in the group. For instance, let p be a prime number. In ${}^{\mathbb{Z}}/_{p^2\mathbb{Z}}$ there is one element of order 1, then $p - 1$ elements of order p and $p^2 - p$ elements of order p^2. The last are precisely the generators for this cyclic group. Instead in the only other group of order p^2, namely ${}^{\mathbb{Z}}/_{p\mathbb{Z}} \times {}^{\mathbb{Z}}/_{p\mathbb{Z}}$, all but 0 have order p.

An abelian group A is called **finitely generated** if there is a finite set $\{a_1, a_2, \dots, a_s\}$ such that every $a \in A$ can be written as a linear combination of these a_i; more precisely, there exists (maybe not unique) integers m_1, m_2, $\dots$, m_s such that $a = m_1\, a_1 + m_2\, a_2 + \cdots + m_s\, a_s$.

Theorem 2.2. *For any finitely generated abelian group A there exists an integer $r \geqslant 0$, called the* ***rank*** *such that $A \cong A_{\mathrm{tors}} \times \mathbb{Z}^r$ where A_{tors} is the finite subgroup of A consisting of all elements of finite order, called the* ***torsion subgroup*** *of A.*

Of course A_{tors} can then be written as a product of cyclic groups as in the previous theorem.

The proofs of these are part of G13GTH= MATH3001.

2.2 Fields

Also from G12ALN=MATH2015, but Section 2.1, we recall the definition of a field:

Definition. A **field** k is a non-empty set endowed with two binary operations $+ : k \times k \to k$, called **addition**, and $\cdot : k \times k \to k$, called **multiplication**, satisfying:

(F1) $\langle k, +\rangle$ is an abelian group. $0 \in k$ denotes the additive identity and $-a$ the additive inverse of $a \in k$.

(F2) $\langle k \setminus \{0\}, \cdot\rangle$ is an abelian group denoted $k^\times$. The multiplicative identity is denoted by $1 \in k$, the inverse of $0 \neq a \in k$ by a^{-1} or $1/a$.

(F3) Distributivity: $a(b + c) = ab + ac$ for all a, b and $c \in k$.

Note we already used the common shorthand ab for the multiplication $a \cdot b$.

Common fields are $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$. The ring ${}^{\mathbb{Z}}\!/_{m\mathbb{Z}}$ of residue classes of integers modulo some integers $m > 1$ was studied a lot in G12ALN=MATH2015. We will simplify the notation and write the elements as 0, 1, $\dots$, $m-1$ when they should be denoted by $[0] = m\mathbb{Z}$, $[1] = 1 + m\mathbb{Z}$, $\dots$, $[m-1] = m - 1 + m\mathbb{Z}$.

Recall from G12ALN=MATH2015 as for instance in Proposition 3.0.10 the following.

Proposition 2.3. *The ring $\mathbb{Z}/n\mathbb{Z}$ is a field if and only if n is a prime number.*

For a prime number p, we denote the field ${}^{\mathbb{Z}}\!/_{p\mathbb{Z}}$ by $\mathbb{F}_p$.

Example. We write out the addition and multiplication tables for $\mathbb{F}_5$.

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

$\diamond$

There are other finite fields, like fields $\mathbb{F}_4$, $\mathbb{F}_8$, $\mathbb{F}_9$, ... with 4, 8, 9 elements, but they are not equal to ${}^{\mathbb{Z}}\!/_{4\mathbb{Z}}$, ...

Further examples of new fields can be obtained as "extensions" of old fields. G13NGA=MATH3021 has plenty of those. For instance the Gaussian numbers $\mathbb{Q}(i)$ which consists of all complex numbers $a + bi$ with a and b in $\mathbb{Q}$ form a field. Similarly the field $\mathbb{Q}(\sqrt{7})$ of all real numbers of the form $a + b\sqrt{7}$ with a and b in $\mathbb{Q}$ is a field. We can illustrate $\mathbb{F}_9$ as elements of the form $a + b\,i$ with a and b in $\mathbb{F}_3$ with the usual rule $i^2 = -1$.

For a field k, we denote by $k[x]$ the ring of polynomials with coefficients in k. Similarly $k[x, y]$ is the ring of polynomials in two variables.

Definition. A field k is called **algebraically closed** if for every polynomial $f \in k[x]$ with coefficients in k, there exists a root of f in k, i.e., $f(\alpha) = 0$ for some $\alpha \in k$.

It follows, by induction, that the only irreducible polynomials with coefficients in an algebraically closed field are of degree 1. The fundamental theorem of algebra states that $\mathbb{C}$ is algebraically closed; the usual proof uses complex function theory. See Chapter 4 in [2]. All other examples above are not algebraically closed. It is true, but not obvious that for any field k, there exists an algebraically closed field $\overline{k}$ containing k.

3 The projective line and plane

3.1 A motivation

We start by motivating the definition of the projective plane by the following diophantine consideration. Suppose we would like to study the rational solutions to the equation

$$x^3 + y^3 = 1. \tag{3.1}$$

We may write a solution (x, y) as fractions of integers. Let us make common denominators: $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$. Now X and Y and Z are integers such that

$$X^3 + Y^3 = Z^3. \tag{3.2}$$

This is a homogeneous equation. If we find a solution (X, Y, Z) to (3.2), like $(1, 0, 1)$, then we find plenty more by multiplying it with any integer λ as $(\lambda \cdot X, \lambda \cdot Y, \lambda \cdot Z)$ is also a solution to (3.2). However they represent the same solution $\left(\frac{X}{Z}, \frac{Y}{Z}\right)$ of the original equation (3.1). We should look for solutions of the homogeneous equation up to multiplication by a scalar as only the ratios between the coordinates matter for the original question.

Another issue is that the second equation has new solutions that were not present in the first. For instance $(1, -1, 0)$ and $(0, 0, 0)$ are solutions to (3.2) which do not give solutions of (3.1) since the denominator Z is not allowed to be zero. We can discard $(0, 0, 0)$ as a trivial solution. At first the other additional solution looks like a problem, but later we will see that the addition of these extra points "at infinity" is a blessing.

3.2 Algebraic definition

Let k be a field. Consider the following relation between vectors of length 3 with coordinates in k.

$$\begin{aligned} &(X, Y, Z) \sim (X', Y', Z') \qquad \text{only if} \\ &\qquad \text{there is a } \lambda \in k^{\times} \text{ such that } X' = \lambda \cdot X \text{ and } Y' = \lambda \cdot Y \text{ and } Z' = \lambda Z. \end{aligned}$$

It was shown in the exercises that this is an equivalence relation.

Definition. The **projective plane** $\mathbb{P}^2$ consists of all equivalent classes of non-zero vectors in k^3. Or in a formula

$$\mathbb{P}^2(k) = \frac{k^3 \setminus \{(0, 0, 0)\}}{\sim}.$$

The elements in $\mathbb{P}^2(k)$ corresponding to the equivalence class of (X, Y, Z) is written as $(X : Y : Z)$. These elements are called the **points** of $\mathbb{P}^2$ with coordinates in k. Of course, it is not yet clear in what sense these are "points" in a "plane".

There is an obvious generalisation to all dimensions:

Definition. The n-dimensional **projective space** $\mathbb{P}^n$ consists of all equivalent classes of non-zero vectors in k^{n+1}. Or in a formula

$$\mathbb{P}^n(k) = \frac{k^{n+1} \setminus \bigl\{(0, 0, \dots, 0)\bigr\}}{\sim}$$

where $(X_0, X_1, \dots, X_n) \sim (\lambda\, X_0, \lambda\, X_1, \dots, \lambda\, X_n)$ for some $\lambda \in k^\times$. The equivalence class of $(X_0, X_1, \dots, X_n)$ is written as $(X_0 : X_1 : \dots : X_n)$.

For instance, we have

$$(1 : 2 : 3) = (2 : 4 : 6) = (-1 : -2 : -3) = (\tfrac{1}{7} : \tfrac{2}{7} : \tfrac{3}{7})$$

for $k = \mathbb{Q}$. The ":" should remind us that we only care about the ratio between the coordinates. For a field like $k = \mathbb{F}_5$ this looks less familiar:

$$(1 : 2) = (2 : 4) = (3 : 1) = (4 : 3),$$

but it works just the same.

Note that $\mathbb{P}^n(k)$ is no longer a vector space. There is not even a good addition defined on it any more. Also there is no meaning of distance in $\mathbb{P}^n(\mathbb{R})$ any more. At this stage, it is really just a set.

3.3 The projective line $\mathbb{P}^1$

Let us look at the case $n = 1$ a bit closer. The points in $\mathbb{P}^1(k)$ are of the form $(X : Y)$ with X and Y in k and at least one is non-zero. If Y is not zero, then $(X : Y) = (\frac{X}{Y} : 1)$. We see that $\mathbb{P}^1(k)$ contains a copy of k by sending x to $(x : 1)$. The only points we are missing are those with $Y = 0$. Now if $(X : 0) \in \mathbb{P}^1(k)$ then X is non-zero and we can divide both coordinates by X to get $(X : 0) = (\frac{X}{X} : \frac{0}{X}) = (1 : 0)$. In other words, there is only one extra point and we have that $\mathbb{P}^1(k)$ is k with one more point or as a formula

$$\mathbb{P}^1(k) = \Bigl\{(x : 1) \Bigm| x \in k\Bigr\} \sqcup \bigl\{(1 : 0)\bigr\} = \text{“}k \sqcup \text{one point”}.$$

In particular $\mathbb{P}^1(\mathbb{F}_p)$ has $p + 1$ points.

Now we will explain why we call the extra point a point "at infinity" in the case $k = \mathbb{R}$. The following sequence of points approaches (in some sense) the point $(1 : 0)$:

$$\begin{aligned}
(1:1) &= (1:1)\\
(1:\tfrac{1}{2}) &= (2:1)\\
(1:\tfrac{1}{3}) &= (3:1)\\
\vdots \quad &\quad \vdots\\
(1:\tfrac{1}{k}) &= (k:1)
\end{aligned}$$

When viewed as points in $\mathbb{R}$ corresponding to $(x:1)$ they tend to infinity. The missing point can be thought of as $\infty = \frac{1}{0}$. However it is better to avoid this and just to write it as $(1:0)$. For instance it is also the point at $-\infty$ as the sequence $(1:-\frac{1}{k}) = (-k:1)$ converges also to $(1:0)$. Figure 5 illustrates this idea. One can consider $\mathbb{P}^1(\mathbb{R})$ a sort of a circle with $(1:0)$ at the top. The rest of the projective line is then identified with a usual line $\mathbb{R}$.

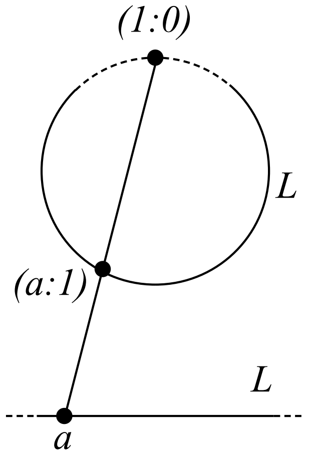


Figure 5: Adding the "point at infinity" $(1:0)$ to a line L in $\mathbb{R}$ creates a circle.

To distinguish, we will call the usual line k now the **affine line** and write it as $\mathbb{A}^1(k)$. From the point of view of the projective line, it was a bit random that we chose to view the points $(a:1)$ as an affine line inside the projective line missing one point $(1:0)$. We could also view $\mathbb{P}^1(k)$ as the affine line $(1:b)$ plus the missing point $(0:1)$. In any case, where ever we are in $\mathbb{P}^1(k)$ it looks close-by very much like $\mathbb{A}^1(k)$ and it makes sense to think of $\mathbb{P}^1(k)$ as a 1-dimensional object.

3.4 The projective plane $\mathbb{P}^2$

Now we are better prepared to have another look at $\mathbb{P}^2(k)$. Again if $Z \neq 0$, then $(X:Y:Z) = (\frac{X}{Z}:\frac{Y}{Z}:1)$. Hence the subset of all $(X:Y:Z)$ with $Z \neq 0$ can be identified with the usual plane, now called the **affine plane** $\mathbb{A}^2(k) = k^2$, via $(x,y) \mapsto (x:y:1)$. This image misses some points at infinity, namely all points of the form $(X:Y:0)$. Since at least one of X or Y is non-zero, we see that the set of points at infinity forms exactly a projective line through the identification $(X:Y) \mapsto (X:Y:0)$.

$$\begin{aligned}
\mathbb{P}^1(k) &= \Big\{(x:y:1) \;\Big|\; (x,y) \in k^2\Big\} \sqcup \Big\{(X:Y:0) \;\Big|\; (X:Y) \in \mathbb{P}^1(k)\Big\}\\
&= \text{“affine plane} \sqcup \text{projective line”}.
\end{aligned}$$

Now to lines within the projective plane. In the affine plane, lines, well affine lines, are described by equations of the form $y = m\,x + n$ where n and m are constants in k; though that already excludes lines of the form $x = n$. Now replacing $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$, we find an equation of the form $Y = m\,X + n\,Z$, that is a linear equation in our three projective variables.

Definition. A (projective) **line in $\mathbb{P}^2$** defined over k is an equation of the form

$$L: \qquad a\,X + b\,Y + c\,Z = 0 \tag{3.3}$$

with constants a, b and c in k such that at least one of them is non-zero.

Lemma 3.1. *Let L be a line defined over k as in equation (3.3). Then the set of solutions $L(k)$ is a well-defined subset of the projective plane $\mathbb{P}^2(k)$.*

Proof. Let X, Y, Z be elements in k such that $(X : Y : Z) \in \mathbb{P}^2(k)$ and $a\,X + b\,Y + c\,Z = 0$. Let $\lambda \in k^\times$. Then $\lambda \cdot X$, $\lambda \cdot Y$ and $\lambda \cdot Z$ also satisfy the equation as $a\,(\lambda X) + b\,(\lambda Y) + c\,(\lambda Z) = 0$. So all elements in the equivalence class at once satisfies the equation and so

$$L(k) = \bigl\{(X : Y : Z) \in \mathbb{P}^2(k) \bigm| a\,X + b\,Y + c\,Z = 0\bigr\}$$

is a well-defined subset of $\mathbb{P}^2(k)$. □

If L is a line in $\mathbb{P}^2$ with $b \neq 0$, then we recover the previous affine line by setting $m = -a/b$ and $n = -c/b$. The affine points in L are those $(x : y : 1)$ in $L(k)$ and this misses out the points at infinity. It is not hard to work out that there is only one point, namely $(b : -a : 0)$ on $L(k)$ at infinity.

If L is a line with $b = 0$, but $a \neq 0$, we recover the affine line $x = -c/a$ together with the point $(0 : 1 : 0)$ at infinity. That only leaves the line $Z = 0$; this is simply the line at infinity itself.

Two equations

$$\begin{aligned} L: &\qquad a\,X + b\,Y + c\,Z = 0 \\ L': &\qquad a'\,X + b'\,Y + c'\,Z = 0 \end{aligned}$$

with constants in k represent the same line if and only if there is a $\lambda \in k^\times$ such that $a' = \lambda \cdot a$ and $b' = \lambda \cdot b$ and $c' = \lambda \cdot c$.

Proposition 3.2. *Any two distinct lines L and L' defined over k intersect in exactly one point in $\mathbb{P}^2(k)$.*

Proof. Consider the matrix

$$M = \begin{pmatrix} a & b & c \\ a' & b' & c' \end{pmatrix}$$

Then we are looking for $(X:Y:Z)$ such that $M \cdot \left(\begin{smallmatrix} X \\ Y \\ Z \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$. Clearly we have some non-zero solution as the number of equations is less than the number of variables. The assumption that L and L' are distinct now implies that the two rows are linearly independent. Therefore the solution space of the system of linear equation is 1-dimensional. In other words, there is a unique equivalent class $(X:Y:Z) \in \mathbb{P}^2(k)$ of non-zero solutions. □

Example. The lines $X + 2\,Y + 3\,Z = 0$ and $4\,X + 5\,Y + 6\,Z = 0$ intersect in the point $(1 : -2 : 1)$ only. ◇

This proposition is a first instance of a statement that is easier in the projective plane than in the affine plane. In the affine plane, two distinct lines are either parallel or they intersect in a single point. So what happens to parallel affine lines? The affine lines $ax + by + c = 0$ and $a'x + b'y + c' = 0$ are parallel exactly when (a, b) is collinear to (a', b'). If so then the intersection is at $(b : -a : 0) = (b' : -a' : 0)$, which is a point at infinity. Conversely, each point $(a : b : 0)$ on the line at infinity is the intersection of a set of parallel affine lines. This is another way to motivate the introduction of the projective plane.

Proposition 3.3. *Through any two distinct points in $\mathbb{P}^2(k)$ passes exactly one projective line.*

Proof. Let $(X:Y:Z)$ and $(X':Y':Z')$ be two distinct points in $\mathbb{P}^2(k)$. We are looking for a, b and c in k such that $M \cdot \left(\begin{smallmatrix} a \\ b \\ c \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right)$ where M is the matrix

$$M = \begin{pmatrix} X & Y & Z \\ X' & Y' & Z' \end{pmatrix}.$$

To say that the two points are distinct is equivalent to the two rows of M being linear independent. Therefore there exists a non-zero solution (a, b, c), unique up to scalar multiplication. In other words, there is a unique projective line $aX + bY + cZ = 0$. □

There is a striking resemblance between the proofs of the two propositions above. Behind this is the principle of duality in projective geometry. In fact there is a bijection between the set of all lines in $\mathbb{P}^2$ and $\mathbb{P}^2$ itself by sending the line L in equation (3.3) to $(a : b : c) \in \mathbb{P}^2(k)$.

Example. The line through $(1:2:3)$ and $(4:5:6)$ is $L: X-2Y+Z=0$. ⋄

Corollary 3.4. *Let $P \in \mathbb{P}^2(k)$ and let L_0 be a line in $\mathbb{P}^2$ defined over k such that $P \notin L_0$. Then there is a bijection between all lines L defined over k passing through P and $L_0(k)$ given by sending L to the intersection of L_0 and L.*

The situation is illustrated in Figure 6. The set of all lines through a point is an example of what algebraic geometers call a "pencil".

Proof. First we use Proposition 3.2: If L is a line through P which is defined over k, then it is distinct from L_0 and hence it intersects in L_0 in exactly one point $Q \in \mathbb{P}^2(k)$. Conversely, if Q is a point in $L_0(k)$, then there is a unique line L defined over k passing through P and Q by Proposition 3.3. □

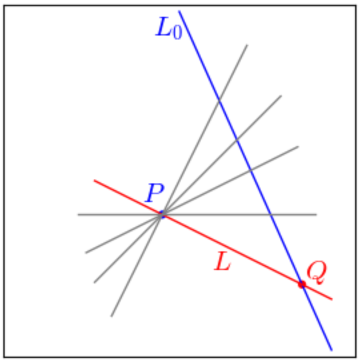


Figure 6: A pencil of lines

3.5 Parametrising lines

With affine lines, we are used to switch between the form where we write them as an equation, like $y = mx + n$, and the parametric form. We would like to have the same at our disposal in the projective plane. For an affine line, we write it as $t \cdot \vec{v} + \vec{w}$ for two constant vectors $\vec{v}$ and $\vec{w}$ in k^2, while t varies through k.

Given two distinct points $P_0 = (X_0 : Y_0 : Z_0)$ and $P_1 = (X_1 : Y_1 : Z_1)$ in $\mathbb{P}^2(k)$, we can consider the points of the form

$$(s\,X_0 + t\,X_1 : s\,Y_0 + t\,Y_1 : s\,Z_0 + t\,Z_1)$$

as $(s:t)$ varies in $\mathbb{P}^1(k)$. When $(s:t) = (1:0)$, we find $P_0 = (X_0 : Y_0 : Z_0)$ and when $(s:t) = (0:1)$, then it is P_1 that we recover. Replacing $(s:t)$ by $(\lambda s : \lambda t)$ for some $\lambda \in k^\times$ will only multiply the above by λ. Therefore the map

$$\ell : \quad (s:t) \mapsto (s\,X_0 + t\,X_1 : s\,Y_0 + t\,Y_1 : s\,Z_0 + t\,Z_1)$$

is a well-defined map from $\mathbb{P}^1(k)$ to a subset of $\mathbb{P}^2(k)$. The image in fact is the unique projective line that passes through P_0 and P_1. We will often just write

$$\ell(s:t) = s\,P_0 + t\,P_1.$$

Example. The line L in the previous example is parametrised by $\ell(s:t) = (s+4t : 2s+5t : 3s+6t)$. We get points like $\ell(1:0) = (1:2:3)$, $\ell(0:1) = (4:5:6)$ or $\ell(1:1) = (5:7:9)$ and so forth. $\diamond$

To recover the equation of the line, we can eliminate the variables s and t in the system

$$\begin{aligned} X &= s\,X_0 + t\,X_1 \\ Y &= s\,Y_0 + t\,Y_1 \\ Z &= s\,Z_0 + t\,Z_1 \end{aligned}$$

It should not be a big surprise to find a formula reminding us of the cross product:

$$L: \qquad (Y_0Z_1 - Y_1Z_0)\cdot X + (Z_0X_1 - Z_1X_0)\cdot Y + (X_0Y_1 - X_1Y_0)\cdot Z = 0.$$

From now on we will often pass from one description to the other. Either as an equation $L : aX + bY + cZ = 0$ with $a, b, c \in k$ or as a map $\ell : \mathbb{P}^1(k) \to L(k)$ as given above with points P_0 and P_1 in $L(k)$. The parametrisation is not unique, but this is the unique parametrisation such that $\ell(1:0) = P_0$ and $\ell(0:1) = P_1$.

Example. For instance in yet the same example as before, we could have used the points $(1:2:3)$ and $(5:7:9)$ as P_0 and P_1 to find another parametrisation $\ell'(s:t) = (s+5t : 2s+7t : 3s+9t)$. This time $\ell'(-1:1) = (4:5:6)$. We can plug in the above formula to get the equation back.

$$L : (2\cdot 9 - 7\cdot 3)X + (3\cdot 5 - 9\cdot 1)Y + (1\cdot 7 - 5\cdot 2)Z = -3(X - 2Y + Z) = 0$$

$\diamond$

4 Projective curves

We start to study curves other than lines in $\mathbb{P}^2$. We still fix a field k. We begin with the easiest case, the quadratic equations. In general our "curves" are always given by polynomial equations. Equations of the form $y = \sin(x)$ involving non-polynomial functions are not considered here.

Homogenisation

Definition. A non-zero polynomial $F \in k[X, Y, Z]$ is **homogeneous** if all monomials appearing have the same constant degree, called the **degree** of F.

For instance $X + Y^2$ is not homogeneous, but $X^3 + X^2Y - 17\,XYZ - Z^3$ is homogeneous.

Lemma 4.1. *The solution set to $F = 0$ for a homogeneous polynomial $F \in k[X, Y, Z]$ is a well-defined subset of $\mathbb{P}^2(k)$.*

Proof. Let $\lambda \in k^\times$. If $i + j + k = d$, then $(\lambda X)^i(\lambda Y)^j(\lambda Z)^k = \lambda^d X^i Y^j Z^k$. We deduce that, for any is homogeneous F of degree d, the formula

$$F(\lambda X, \lambda Y, \lambda Z) = \lambda^d \cdot F(X, Y, Z)$$

holds because this is true for any monomial of degree d. It follows that if (X, Y, Z) is a solution to $F = 0$ then so is $(\lambda X, \lambda Y, \lambda Z)$. □

We illustrate by an example how we get from an affine equation to a projective one. If the equation is

$$x^4 - 5x^2y^2 + 4y^4 + 5x^2 + 9y + 10 = 0. \tag{4.1}$$

Then we write x as $\frac{X}{Z}$ and y as $\frac{Y}{Z}$. We get

$$\frac{X^4}{Z^4} - 5\frac{X^2Y^2}{Z^4} + 4\frac{Y^4}{Z^4} + 5\frac{X^2}{Z^2} + 9\frac{Y}{Z} + 10 = 0.$$

Now we multiply through by Z^4 to obtain

$$X^4 - 5X^2Y^2 + 4Y^4 + 5X^2Z^2 + 9YZ^3 + 10Z^4 = 0.$$

This is a homogeneous polynomial of degree 4. Of course, we get the same result by replacing x by X and y by Y and then multiply each monomial by the appropriate power of Z to have a homogeneous polynomial.

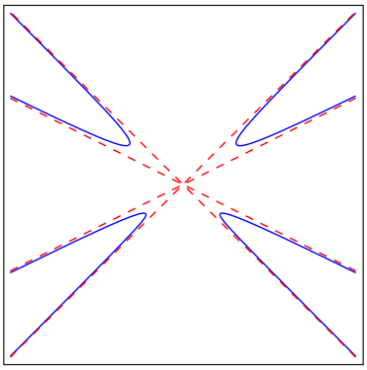


Figure 7: A picture of $C(\mathbb{R})$ for the example (4.1)

Definition. In general if $f \in k[x,y]$ we define the **homogenisation** of f in the same way: First, we replace all x by X/Z and all y by Y/Z, then we multiply through by the smallest power of Z needed to clear the Z appearing in denominators.

Explicitly, let

$$f = \sum_{i=0}^{u} \sum_{j=0}^{v} c_{i,j}\, x^i y^j$$

with $c_{i,j} \in k$. The degree d of f is the largest $i+j$ such that $c_{i,j} \neq 0$. Then

$$F = \sum_{i=0}^{u} \sum_{j=0}^{v} c_{i,j}\, X^i Y^j Z^{d-i-j} \in k[X,Y,Z]$$

is the homogenisation of f.

Definition. If f is a non-constant polynomial in $k[x,y]$, we call $C : f = 0$ an **affine curve** (defined over k). If $F \in k[X,Y,Z]$ is the homogenisation of f then we call $\overline{C} : F = 0$ the **projective closure** of C. It is a **projective curve** defined over k.

As before, we will write in this case $C(k)$ for the solution set to $f = 0$, which is a subset of k^2. Similarly $\overline{C}(k)$ will denote the solution set to $F = 0$ as a subset of $\mathbb{P}^2(k)$; this makes sense by the above lemma. We have a disjoint union $\overline{C}(k) = C(k) \sqcup \bigl(\overline{C} \cap L_\infty\bigr)(k)$ where $L_\infty : Z = 0$ is the line at infinity.

In the above example (4.1), say with $k = \mathbb{R}$,

$$\begin{aligned} C : &\quad x^4 - 5x^2y^2 + 4y^4 + 5x^2 + 9y + 10 = 0 \\ \overline{C} : &\quad X^4 - 5X^2Y^2 + 4Y^4 + 5X^2Z^2 + 9YZ^3 + 10Z^4 = 0 \end{aligned}$$

The intersection with L_∞ is given by $(X : Y : 0)$ with

$$X^4 - 5X^2Y^2 + 4Y^4 = (X + 2Y)(X + Y)(X - Y)(X - 2Y) = 0$$

so exactly by the four points $\bigl\{(-2 : 1 : 0), (-1 : 1 : 0), (1 : 1 : 0), (2 : 1 : 0)\bigr\}$. One can think of these four points as the four asymptotes of C given by $x = \pm 2y$ and $x = \pm y$. See Figure 7 for a real picture of this curve.

We make a short excursion to talk about homogeneous polynomials in two variable. Let $G(S, T) \in k[S, T]$ be a homogeneous polynomial of degree d, say

$$G(S, T) = a_0 S^d + a_1 S^{d-1} T + \cdots + a_{d-1} S T^{d-1} + a_d T^d.$$

Write $g(x) = G(x, 1) \in k[x]$ for the associated polynomial in one variable. This can now be factored into irreducible factors as studies in Section 2 of G12ALN=MATH2015. Some of these factors are linear corresponding to solutions of $g(x)$ in k. Putting all the other factors into one polynomial $h(x)$, we can write

$$g(x) = (x - x_1)^{m_1} \cdot (x - x_2)^{m_2} \cdots (x - x_e)^{m_e} \cdot h(x).$$

The homogenisation brings us to

$$G(S, T) = g\Bigl(\frac{S}{T}\Bigr)\, T^d = (S - x_1\, T)^{m_1} \cdot (S - x_2\, T)^{m_2} \cdots (S - x_e\, T)^{m_e} \cdot \tilde{H}(S, T).$$

For a homogeneous polynomial $\tilde{H} \in k[S, T]$. If $a_0 = 0$ then T will divide $\tilde{H}(S, T)$, but other than that H cannot have any linear factors any more. We can factor out this power of T from $\tilde{H}$ to get a homogeneous polynomial H with no linear factors any more. Thinking of $(s_i : t_i) = (x_i : 1)$ as points on the projective line $\mathbb{P}^1$ the extra linear factors T will correspond to the point $(1 : 0)$ at infinity.

Lemma 4.2. *Any homogeneous polynomial $G \in k[S, T]$ of degree $d > 0$ can be written as*

$$G(S, T) = (t_1\, S - s_1 T)^{m_1} \cdot (t_2\, S - s_2\, T)^{m_2} \cdots (t_e\, S - s_e\, T)^{m_e} \cdot H(S, T)$$

for a homogeneous polynomial $H \in k[S, T]$ of degree $d - m_1 - m_2 \cdots - m_e$ and where $(s_1 : t_1)$, $(s_2 : t_2)$, ..., $(s_e : t_e)$ are precisely all points $(s : t)$ in $\mathbb{P}^1(k)$ such that $G(s : t) = 0$.

It is also to note that $H = 1$ if k is algebraically closed as $g(x)$ must then factor completely into linear factors.

Example. Consider $G(S, T) = S^2T^2 - 6ST^3 + 13T^4$. If $k = \mathbb{Q}$, this factors as $G(S, T) = T^2 \cdot (S^2 - 6ST + 13T^2)$. Instead for $k = \mathbb{C}$ the last factor splits up further as $G(S, T) = T^2 \cdot (S - (3 + 2i)T) \cdot (S - (3 - 2i)T)$. ◇

4.1 Conics

A conic is an equation defined by a homogeneous polynomial of degree 2; more explicitly:

Definition. A **conic** in $\mathbb{P}^2$ defined over k is an equation of the form

$$C: \qquad a_0\,X^2 + a_1\,XY + a_2\,XZ + a_3\,Y^2 + a_4\,YZ + a_5\,Z^2 = 0 \tag{4.2}$$

where $a_0, \ldots, a_5$ are constants in k, which are not all equal to zero.

Example. The projective closure of the (affine) circle $x^2 + y^2 = r$ is the conic $X^2 + Y^2 - rZ^2 = 0$. The hyperbola $xy = c$ yields the conic $XY - cZ^2 = 0$. The parabola $y = x^2$ gives the conic $-X^2 + YZ = 0$. The conic $XY = 0$ really looks like the union of the two axis. ◇

Definition. A conic $F = 0$ is **degenerate** if the quadratic polynomial F is the product of two linear polynomials over some field K containing k. Otherwise the conic is **non-degenerate**.

The solution set $C(K)$ of a degenerate conic is the union of two (not necessarily distinct) lines, called the **components** of the degenerate conic.

Note that K may be different than k. For instance the conic $C : X^2 + 2XY + Y^2 + Z^2 = 0$ for $k = \mathbb{Q}$ does not factor into two linear polynomials with coefficients in $\mathbb{Q}$, but over $\mathbb{C}$ one has

$$0 = X^2 + 2XY + Y^2 + Z^2 = (X + Y + iZ)(X + Y - iZ).$$

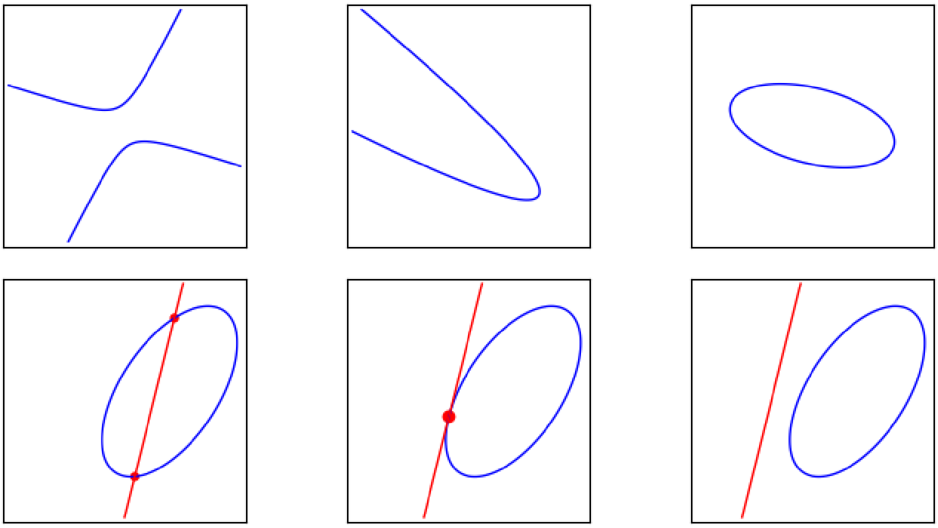
Example. Let C be a non-degenerate conic defined by equation (4.2) over $k = \mathbb{R}$. Let us consider the intersection of C with the line $Z = 0$ at infinity. The intersection are points $(X : Y : 0)$ with $a_0X^2 + a_1XY + a_3Y^2 = 0$. If all three coefficients were zero, the conic would be degenerate. Consider $\Delta = a_1^2 - 4a_0a_3$. The quadratic polynomial $a_0X^2 + a_1XY + a_3Y^2$ factors into two linear terms exactly when $\Delta \geqslant 0$. There are now three cases, illustrated in Figure 8.

- If $\Delta > 0$, then there are two distinct solutions in $\mathbb{R}$ and hence two distinct points $(X : Y : 0)$ in $C(\mathbb{R})$. The case $a_0 = 0$ and $a_1 \neq 0$ is included here. The affine part of the conic is a hyperbola and the two points at infinity correspond to the two asymptotes.

- If $\Delta = 0$, there is exactly one solution in $\mathbb{R}$ and hence one point in $C(\mathbb{R})$ at infinity. This also includes the case $a_0 = a_1 = 0$. The affine part is a parabola with the unique point at infinity indicating the axis of symmetry. One should think of the line at infinity as the tangent to C at this point.

- If $\Delta < 0$, there is no solution in $\mathbb{R}$ and hence $C \cap \{Z = 0\} = \varnothing$. In this case the affine piece of C is an ellipse. This includes the case of an empty ellipse, like $X^2 + Y^2 + Z^2 = 0$.

It now turns out that theory of projective conics explains the three types of real "conic sections" in a coherent and aesthetically pleasing manner. ◇



(a) $\Delta > 0$ (b) $\Delta = 0$ (c) $\Delta < 0$

Figure 8: The three types of affine pieces of a real conic (top row) and the positioning of the conic and the line at infinity (bottom row)

Lemma 4.3. *Let C be a conic and let L be a line, both defined over k. Then the intersection contains at most 2 points unless the conic is degenerate and L is one of its components.*

Example. We illustrate the lemma with an example over different fields k. The conic $C : X^2 + Y^2 - 8\,Z^2 = 0$ has an affine picture of a circle. Intersecting with the line $L : X - Y - 2Z = 0$ can be done using a parametrisation $\ell(s : t) = (2s : 2t : s - t)$. Plugging in the equation of C we obtain

$$(2s)^2 + (2t)^2 - 8(s - t)^2 = 0$$

which simplifies to $s^2 - 4st + t^2 = 0$. If the field is $k = \mathbb{Q}$ then there are no solution to this equation (as the discriminant is 12) and so $C(\mathbb{Q}) \cap L(\mathbb{Q}) = \varnothing$.

Instead for $k = \mathbb{R}$, the polynomial factors as $(s - (2+\sqrt{3})t)(s - (2-\sqrt{3})$. Then we can find two solutions $\ell(2 \pm \sqrt{3} : 1) = (4 \pm 2\sqrt{3} : 2 : 1 \pm \sqrt{3}) = (1 \pm \sqrt{3} : -1 \pm \sqrt{3} : 1)$. Similar for $k = \mathbb{F}_{11}$, we find the two points $(6 : 4 : 1)$ and $(7 : 5 : 1)$.

The case $k = \mathbb{F}_3$ is special. Since the equation factors as $(s+t)^2 = 0$, there is only one point $\ell(1 : -1) = (1 : 2 : 1)$. We will later see that we should count this point twice in some sense as L turns out to be the tangent to C. ◇

Proof. We know that we can parametrise the line by a map

$$\ell(s : t) = (\alpha s + \alpha' t : \beta s + \beta' t : \gamma s + \gamma' t)$$

for two points $P = (\alpha : \beta : \gamma)$ and $Q = (\alpha' : \beta' : \gamma')$ in $\mathbb{P}^2(k)$. Substitute this into the equation (4.2) of the conic:

$$0 = a_0(\alpha s + \alpha' t)^2 + a_1(\alpha s + \alpha' t)(\beta s + \beta' t) + \cdots + a_5(\gamma s + \gamma' t)^2.$$

Rearranging this equation one obtains an equation of the form

$$0 = c_0\, s^2 + c_1\, st + c_2\, t^2 \tag{4.3}$$

with constants c_0, c_1, $c_2 \in k$ depending on P and Q. It could happen that all three constants are zero.

In the case that at least one coefficient is non-zero then we are in the situation of Lemma 4.2. Hence there are at most two points in the intersection of C and L.

We are left with the case when $c_0 = c_1 = c_2 = 0$. This means that every point on L also belongs to C. Without loss of generality, we may assume that c in the equation $L : aX + bY + cZ = 0$ is non-zero. Write $Z' = aX + bY + cZ$. Solve this on Z and substitute this back into the equation (4.2). We find an equation of the form

$$a_0' X^2 + a_1' XY + a_2' XZ' + a_3' Y^2 + a_4' YZ' + a_5' Z'^2 = 0$$

for some constants $a_0', \ldots, a_5' \in k$. Now, all the points with $Z' = 0$ are solution to this new quadratic equation. This implies that $a_0' X^2 + a_1' XY + a_3' Y^2$ is zero for all values of X and Y. Taking $X = 0$ and $Y = 1$, we deduce that $a_3' = 0$. The other way around gives $a_0' = 0$. Finally $X = Y = 1$ concludes $a_1' = 0$. We have obtained that

$$a_0 X^2 + a_1 XY + a_2 XZ + a_3 Y^2 + a_4 YZ + a_5 Z^2 = Z' \cdot (a_2' X + a_4' Y + a_5' Z)$$

and therefore C is degenerate and the line $L : Z' = aX + bY + cZ = 0$ is a component. □

Theorem 4.4. *Let C be a non-degenerate conic defined over k. Either $C(k)$ is empty or there is a bijection between $C(k)$ and $\mathbb{P}^1(k)$.*

Examples for empty $C(k)$ are easy to find: For instance $X^2 + Y^2 + Z^2 = 0$ and $k = \mathbb{R}$. Here are a two additional statements that we will prove in the exercises: The set $C(k)$ is never empty when k is algebraically closed, like for $k = \mathbb{C}$. The same is true when k is a finite field, like $k = \mathbb{F}_p$. In particular, for any prime p and any non-degenerate conic C defined over $\mathbb{F}_p$ there are exactly $p + 1$ solutions in $C(\mathbb{F}_p)$.

Start of proof. We suppose that $C(k)$ is not empty and we wish to construct a bijection using one starting point $P \in C(k)$. Let $\mathfrak{X}$ be the set of all lines in $\mathbb{P}^2$ defined over k that pass through P. By Corollary 3.4, we have a bijection between $\mathbb{P}^1(k)$ and $\mathfrak{X}$. We are left to show that there is a bijection between lines L in $\mathfrak{X}$ and points in $C(k)$.

In the first direction: If Q is a point different from P in $C(k)$, then there is a unique line L defined over k through P and Q. This is due to Proposition 3.3. This gives a map $\phi\colon C(k) \setminus \{P\} \to \mathfrak{X}$.

In the other direction: Let $L \in \mathfrak{X}$. Since C is non-degenerate, L cannot be a component of C. By the previous lemma, the line L intersects C in at most two points in $\mathbb{P}^2(k)$. Let us have a closer look at the proof of Lemma 4.3. We may suppose that $\ell(1 : 0) = P$, which means $P = (\alpha : \beta : \gamma)$. Now, we know that P belongs to the intersection of L and C. Therefore the equation (4.3) has already one solution in k when $(s : t) = (1 : 0)$, therefore it must factor as

$$0 = c_0\, s^2 + c_1\, st + c_2\, t^2 = t\,(c_1\, s + c_2\, t).$$

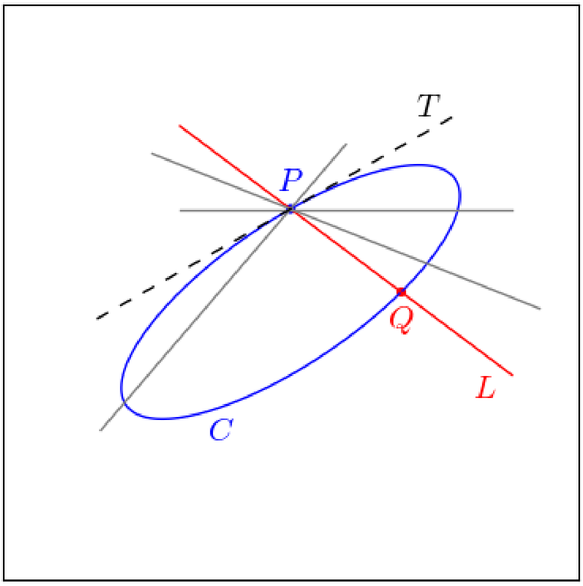


Figure 9: Parametrising a conic with a point

We conclude that there is exactly one more point corresponding to $(s : t) =$

$(-c_2 : c_1)$ in $(C \cap L)(k)$ except in the case when $c_1 = 0$. Therefore ϕ sets up a bijection between $C(k) \setminus \{P\}$ and the set of lines L with $c_1 \neq 0$.

Example. We interrupt the proof to illustrate what we found so far in an example. Let $C : XY - Z^2 = 0$ be the conic defined over $\mathbb{Q}$. We can pick the point $P = (1 : 1 : 1)$ on $C(\mathbb{Q})$. To find the pencil of all lines through P, we can vary a point $R = (a : 0 : b)$ on the x-axis where $(a : b) \in \mathbb{P}^1(\mathbb{Q})$. The line L through P and R is parametrised by $\ell(s : t) = (s + at : s : s + bt)$. We plug this into the equation for C and find

$$0 = (s + at)s - (s + bt)^2 = s^2 + ast - s^2 - 2bst - b^2t^2 = t\big((a - 2b)s - b^2t^2\big).$$

The solution $(s : t) = (1 : 0)$ corresponds to P which belongs to the intersection of C and L. The second point of intersection is

$$\begin{aligned} Q = \ell(b^2 : a - 2b) &= \big(b^2 + a(a - 2b^2) : b^2 : b^2 + b(a - 2b)\big) \\ &= \big(a^2 - 2ab + b^2 : b^2 : ab - b^2\big). \end{aligned}$$

As $(a : b)$ varies through $\mathbb{P}^1(\mathbb{Q})$ the point Q travels through the points in $C(\mathbb{Q})$.

Wait, what was the problem with $c_1 = 0$? In our case this is when $a-2b = 0$, that is to say for a unique point $(a : b) = (2 : 1)$. In that situation the line L intersects C only at the point Q. Luckily when you put that point into the formula for Q anyway, it just returns $P = (1 : 1 : 1)$. So in our case this map $(a : b) \mapsto (a^2 - 2ab + b^2 : b^2 : ab - b^2)$ really sets up a bijection between $\mathbb{P}^1(\mathbb{Q})$ and $C(\mathbb{Q})$. ◇

We return to the proof. Recall that so far we have found that ϕ gives a bijection between $C(k) \setminus \{P\}$ and the set of lines L with $c_1 \neq 0$. It remains to show that there is exactly one line T with $c_1 = 0$ as then we may extend ϕ to a bijection from $C(k)$ to $\mathfrak{X}$ by sending P to T. From the picture in Figure 9, we already guess that this line is the unique tangent T at P to the curve C.

Expanding out and regrouping (yes that is a bit of work, but you should do it once) we find that $c_1 = 0$ is the same as asking that $Q = (\alpha' : \beta' : \gamma')$ is a point satisfying the equation

$$(2a_0\alpha + a_1\beta + a_2\gamma)\, X + (a_1\alpha + 2a_3\beta + a_4\gamma)\, Y + (a_2\alpha + a_4\beta + 2a_5\gamma)\, Z = 0. \quad (4.4)$$

This is indeed the equation of a line passing through P unless all the coefficients above are zero. If they were all zero then any point Q and hence any line L would have $c_1 = 0$.

In other words, we are left with excluding the possibility that $c_1 = 0$ for all lines $L \in \mathfrak{X}$. We will later in Section 4.3 see a proof that this situation implies that C is degenerate. Therefore under our assumption this case does not arise. ⊠

Note that the earlier example $X^2 + 2XY + Y^2 + Z^2 = 0$ is a degenerate example with $C(\mathbb{R}) = \{(1 : -1 : 0)\}$ containing a single point. Here $c_1 = 0$ for all lines.

4.2 Multiplicities and singularities

We return to the study of general projective curves. The last section highlighted that we should be interested in the concept of a tangent and we should understand better the intersection of lines and curves. The idea of the counting certain intersections of lines and curves with multiplicity is well illustrated by Figure 10. As the chord approaches the tangent, two intersection points merge into one point.

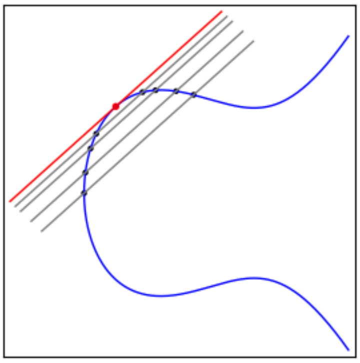


Figure 10: As a chord becomes a tangent, two points merge

Definition. Let C be a projective curve given by an equation $F = 0$ for a non-zero homogeneous polynomial $F \in k[X, Y, Z]$ of degree d. Let $P \in C(k)$ and let L be a line defined over k passing through P. If $\ell(s : t)$ parametrises the line such that $\ell(1 : 0) = P$, then we define the **intersection multiplicity** of C and L at P to be the integer m such that

$$F\bigl(\ell(s:t)\bigr) = t^m \cdot \bigl(a_m\, s^{d-m} + a_{m+1}\, s^{d-m-1} t + \cdots + a_d\, t^{d-m}\bigr)$$

with $a_m \neq 0$. We write $I_P(C, L) = m$.

The hypothesis $P \in C(k) \cap L(k)$ implies that $F(\ell(1 : 0)) = 0$. In other words $I_P(C, L) = m > 0$. If P is not a point of intersection of L and C one defines $I_P(C, L) = 0$.

Note that the homogeneous variable s is not really needed; here is the affine version. Let C be the affine curve given by $f = 0$ for a non-zero polynomial $f \in k[x, y]$. Suppose $P = (0, 0)$ is a point in $C(k)$. We could always change the affine coordinates to move the point to the origin. Let $L : a\,y = b\,x$ be a line through P, which we may parametrise by $t \mapsto (a\,t, b\,t)$. Then

$$f(at, bt) = a_m\, t^m + a_{m+1}\, t^{m+1} + \cdots \in k[t]$$

with $a_m \neq 0$ and $m = I_P(\bar{C}, \bar{L})$ is the intersection multiplicity.

Example. Consider the conic $C : X^2 + Y^2 - YZ = 0$. See Figure 11 for a picture. First we intersect it with the "y-axis", that is with the line $L_1 : X = 0$. There are two points of intersection $P_1 = (0 : 1 : 1)$ and $P_2 = (0 : 0 : 1)$. For the first point P_1, we use the parametrisation $\ell(s : t) = (0 : s : s + t)$. Then $\ell(1 : 0) = P_1$. We obtain

$$F\bigl(\ell(s : t)\bigr) = 0^2 + s^2 - s(s + t) = -st$$

Therefore $I_{P_1}(C, L_1) = 1$. For the second point P_2, we use the parametrisation $\ell(s : t) = (0 : t : s)$; this gives $F\bigl(\ell(s : t)\bigr) = 0^2 + t^2 - ts$. Once again, $I_{P_2}(C, L_1) = 1$.

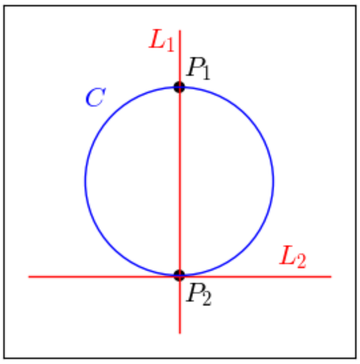


Figure 11: Intersection multiplicities

If we consider the line $L : Y = 0$ instead. There is only one point of intersection namely $P_2 = (0 : 0 : 1)$. We can use the parametrisation $\ell(s : t) = (t : 0 : s)$. This time, we find $F\bigl(\ell(s : t)\bigr) = t^2$ and hence $I_{P_2}(C, L_2) = 2$.

Note that in the above we had a choice for ℓ. For instance in the last example, we could take $\ell(s : t) = (7t : 0 : s + t)$. The result should be the same as we show in the next lemma. ◇

Lemma 4.5. *The definition of $I_P(C, L)$ does not depend on the choice of the parametrisation ℓ of L.*

Proof. If $\ell(s : t) = sP + tQ$ is our first parametrisation, then any other parametrisation ℓ' with $\ell'(1 : 0) = P$ of the same line is of the form $\ell'(s : t) = sP + tQ'$. Since $Q' = aP + bQ$ for some a and $b \neq 0$ in k, we can write

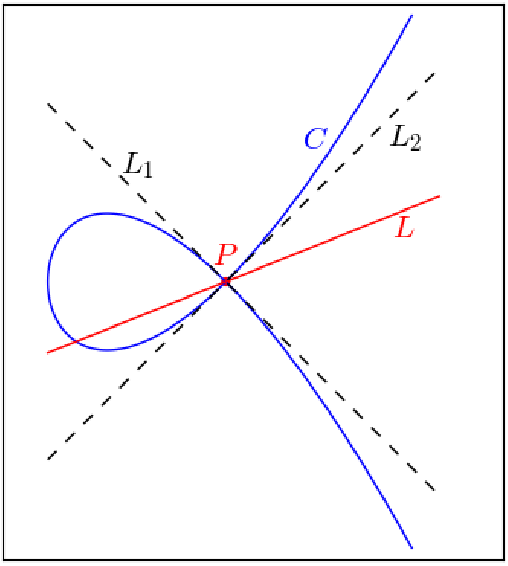


Figure 12: Multiplicity at a singular point

$\ell'(s:t) = (s+at)P + btQ = \ell(s+at : bt)$. Now the terms with the lowest power of t look like

$$F\bigl(\ell'(s:t)\bigr) = F\bigl(\ell(s+at:bt)\bigr) = a_m(s+at)^{d-m}(bt)^m + \cdots = a_m b^m s^{d-m} t^m + \cdots$$

and hence $b \neq 0$ shows that the multiplicity is the same. □

Example. Consider the cubic curve $C : X^3 + X^2Z - Y^2Z = 0$ at the point $P = (0:0:1)$. See Figure 12. Pick two elements a, b in k not both zero and consider the line $L : bX - aY = 0$ going through P; any line through P can be written like this for some $(a:b) \in \mathbb{P}^1(k)$. This line can be parametrised by $\ell(s:t) = (at:bt:s)$. We find

$$F(\ell(s:t)) = (at)^3 + (at)^2 s - (bt)^2 s = (a^2 - b^2)st^2 + a^3t^3.$$

If $a^2 - b^2 \neq 0$, then $I_P(C,L) = 2$. Otherwise when $a = b$, that is the line $L_1 : Y = X$, or when $a = -b$, that is the line $L_2 : Y = -X$, then $I_P(C, L_1) = I_P(C, L_2) = 3$. ◇

This last example should motivated the following definition.

Definition. If a point P on a projective curve C defined over k satisfies the following condition:

For every line L defined over k passing through P, we have $I_P(C,L) > 1$

then we say that P is a **singular point** on C. Otherwise, the point is called a **non-singular point**.

In order to obtain a useful criterion for a point to be singular, we need "derivatives". If $k = \mathbb{R}$ or $k = \mathbb{C}$, we could just use the usual definition involving limits, but for fields like $k = \mathbb{F}_p$ this is not possible. But we may define it formally on polynomials

Definition. If $F = \sum_{i,j} c_{i,j}\, X^i Y^j Z^{d-i-j}$ is a homogeneous polynomial of degree d with coefficients in k, then we define the **partial derivatives** by the formula

$$\frac{\partial F}{\partial X} = \sum_{i,j} i \cdot c_{i,j}\, X^{i-1} Y^j Z^{d-i-j}$$

The partial derivatives $\partial F/\partial Y$ and $\partial F/\partial Z$ are defined similarly. Of course, this extends also to polynomials in more or less variables.

The partial derivative $\partial F/\partial X$ is a homogeneous polynomial with coefficients in k of degree $d-1$ or less. Yes, it could be less. For instance in $k = \mathbb{F}_7$ and $F = X^7$, we have $\partial F/\partial X = 0$. We won't do it, but one can check that all the usual rules of derivations still hold. $\partial(F+G)/\partial X = \partial F/\partial X + \partial G/\partial X$, Leibniz' rule $\partial(F \cdot G)/\partial X = F \cdot \partial G/\partial X + G \cdot \partial F/\partial X$ and the chain rule are all ok.

Theorem 4.6. *Let $C : F = 0$ be a projective curve defined over k given by a homogeneous polynomial $F \in k[X,Y,Z]$ of degree d and let $P \in C(k)$. Then the following three statements are equivalent:*

(a). P is a singular point, i.e., for all lines L defined over k passing through P, we have $I_P(C,L) > 1$.

(b). There are two distinct lines L_1 and L_2 defined over k passing through P such that $I_P(C,L_1) > 1$ and $I_P(C,L_2) > 1$.

(c). $\frac{\partial F}{\partial X}(P) = \frac{\partial F}{\partial Y}(P) = \frac{\partial F}{\partial Z}(P) = 0$.

Example. Let us illustrate the theorem in the previous example pictured in Figure 12. The point is $P = (0 : 0 : 1)$ on $C : X^3 + X^2Z - Y^2Z = 0$. We have seen that all lines L passing through P have intersection multiplicity 2 or 3 at P. Thus the two first points in the theorem are satisfied. The partial derivatives

$$\frac{\partial F}{\partial X} = 3X^2 + 2XZ \qquad \frac{\partial F}{\partial Y} = -2YZ \qquad \frac{\partial F}{\partial Z} = X^2 - Y^2$$

all vanish at $P = (0:0:1)$. ◇

Proof. First we note the following: If $\ell(s:t)$ parametrises L and $F\bigl(\ell(1:t)\bigr) = a_1\, t + a_2\, t^2 + \cdots + a_d\, t^d$ then a_1 is the value of $\partial F\bigl(\ell(1:t)\bigr)/\partial t$ at $t=0$. Therefore $I_P(C,L) > 1$ is equivalent to

$$\left.\frac{\partial F\bigl(\ell(1:t)\bigr)}{\partial t}\right|_{t=0} = 0.$$

We start by showing that (c)⇒(a). Let L be any line passing through P. Pick a point $Q = (X_1 : Y_1 : Z_1)$ on L different from P in order to have a parametrisation $\ell(s:t) = s\,P + t\,Q$ of L. We compute

$$\begin{aligned}\left.\frac{\partial F\bigl(\ell(1:t)\bigr)}{\partial t}\right|_{t=0} &= \left.\frac{\partial F(P+tQ)}{\partial t}\right|_{t=0}\\ &= \frac{\partial F}{\partial X}(P)\cdot X_1 + \frac{\partial F}{\partial Y}(P)\cdot Y_1 + \frac{\partial F}{\partial Z}(P)\cdot Z_1.\end{aligned}$$

Now by the assumption in (c), the last expression is zero and this implies that $I_P(C,L) > 1$ by the remark at the start of the proof.

As the implication (a)⇒(b) is clear, we pass to the proof that (b)⇒(c). Pick two points $Q_1 = (X_1 : Y_1 : Z_1)$ and $Q_2 = (X_2 : Y_2 : Z_2)$, the first on L_1 and the second on L_2 but both different from $P = (X_0 : Y_0 : Z_0)$. By hypothesis we know that $\frac{\partial F(P+tQ_i)}{\partial t}\bigr|_{t=0} = 0$ and by the computation above this gives the two linear equations

$$\begin{aligned}\frac{\partial F}{\partial X}(P)\cdot X_1 + \frac{\partial F}{\partial Y}(P)\cdot Y_1 + \frac{\partial F}{\partial Z}(P)\cdot Z_1 = 0;\\ \frac{\partial F}{\partial X}(P)\cdot X_2 + \frac{\partial F}{\partial Y}(P)\cdot Y_2 + \frac{\partial F}{\partial Z}(P)\cdot Z_2 = 0.\end{aligned}$$

Here is how we obtain a third linear equation. For any t, we have $F(tX_0 : tY_0 : tZ_0) = 0$ as $P \in C(k)$. Taking the derivative of this with respect to t at $t=0$ yields

$$0 = \left.\frac{\partial F(tP)}{\partial t}\right|_{t=0} = \frac{\partial F}{\partial X}(P)\cdot X_0 + \frac{\partial F}{\partial Y}(P)\cdot Y_0 + \frac{\partial F}{\partial Z}(P)\cdot Z_0. \tag{4.5}$$

We gather that the vector $\vec{v} = (\partial F/\partial X(P), \partial F/\partial Y(P), \partial F/\partial Z(P))$ is a solution to the equation $M\vec{v} = \vec{0}$ with

$$M = \begin{pmatrix} X_0 & Y_0 & Z_0 \\ X_1 & Y_1 & Z_1 \\ X_2 & Y_2 & Z_2 \end{pmatrix}.$$

This matrix is non-singular as P and Q_1 and Q_3 do not lie on one line. Hence $\vec{v} = \vec{0}$ and the condition in (c) follows. □

From the negation of condition (b) we deduce the following: If P is a non-singular point on a projective curve C, then there is at most one line L defined over k with $I_P(C, L) > 1$. In fact the proof of the theorem shows us that the equation of this one line is necessarily

$$T_P(C): \qquad \frac{\partial F}{\partial X}(P)\cdot X + \frac{\partial F}{\partial Y}(P)\cdot Y + \frac{\partial F}{\partial Z}(P)\cdot Z = 0 \tag{4.6}$$

Let $C : X^3 - 9\,XZ^2 + 9\,Z^3 - Y^2Z = 0$ for $k = \mathbb{Q}$, illustrated in Figure 13. The derivatives are

$$\frac{\partial F}{\partial X} = 3X^2 - 9Z^2$$
$$\frac{\partial F}{\partial Y} = -2YZ$$
$$\frac{\partial F}{\partial Z} = -18XZ + 27Z^2 - Y^2$$

First for the point $P = (-3 : 3 : 1) \in C(k)$. We find that all three above functions are non-zero at P, therefore P is a non-singular point and the tangent is given by

$$T_P(C): \qquad 18X - 6Y + 72Z = 0$$

We can check that this equation really passes through P confirming (4.5). One can parametrise $T_P(C)$ by $\ell(s:t) = (-3s : 3s + 12t : s + t)$. A tedious computation will reveal $F\bigl(\ell(s:t)\bigr) = -162st^2 - 135t^3$ and hence $I_P\bigl(C, T_P(C)\bigr) = 2$.

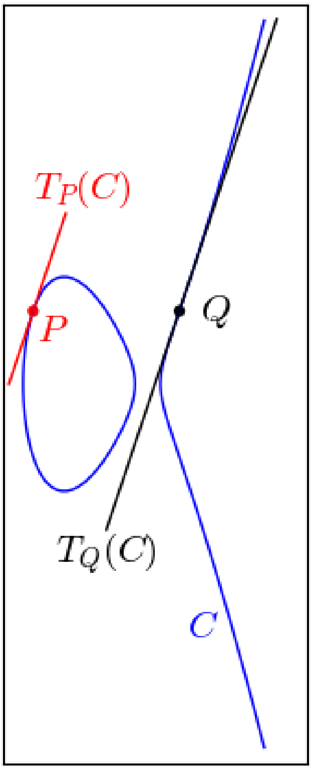


Figure 13: Two tangents

Instead, we can also consider the point $Q = (3 : 3 : 1)$. Again it is a non-singular point and the tangent line is given by

$$T_Q(C) : \qquad 18X - 6Y - 36Z = 0$$

which may be parametrised by $\ell(s : t) = (3s : 3s - 6t : s + t)$. Then $F\bigl(\ell(s : t)\bigr) = -27\,t^3$ shows that $I_Q\bigl(C, T_Q(C)\bigr) = 3$.

This line is called the **tangent line** to C at P, denoted by $T_P(C)$. Sometimes one can also defined one or several tangent lines at a singular point, like in Figure 12; the tangents are then the lines which have higher intersection than all the other lines.

Definition. A non-singular point P on a projective curve C is called an **inflection point** if $I_P\bigl(C, T_P(C)\bigr) > 2$.

So the point Q above is an inflection point and the picture in Figure 13 shows that the tangent at Q passes "through" the curve C.

Example. We return to the example above $C : X^3 - 9\,XZ^2 + 9\,Z^3 - Y^2Z = 0$ for $k = \mathbb{Q}$. This time we look at the point $R = (0 : 1 : 0)$ which can be checked to be the only point on C lying on the line at infinity $Z = 0$. This time the tangent is $Z = 0$ which we can parametrise by $\ell(t : s : 0)$ and $F\bigl(\ell(s : t)\bigr) = t^3$ shows that this point is also an inflection point on C.

Finally to conclude this example, we wish to determine if there are any singular points in $C(\mathbb{C})$? Suppose $P = (X : Y : Z)$ is a singular point then the three partial derivatives must vanish. In particular we have $-2YZ = 0$. Now if $Z = 0$, then we fall back onto the point at infinity that we have just discussed. Therefore we may assume that $Y = 0$. The equation $\partial F/\partial X(P) = 0$ gives $X^3 = 3Z^3$. Therefore the only two singular points could possibly be $\bigl(\sqrt{3} : 0 : 1\bigr)$ and $\bigl(-\sqrt{3} : 0 : 1\bigr)$. However these do not satisfy the equation for C. Therefore C has no singular points defined over $\mathbb{C}$.

The situation is the same for any field K of characteristic 0. However it is a different thing for $k = \mathbb{F}_3$ and $\mathbb{F}_2$. For the first the point $(0 : 0 : 1)$ and for the second the point $(1 : 1 : 1)$ are singular. If the characteristic is different from 2 and 3 then there are no singular points on C. ◇

Definition. Let C be a projective curve defined over k. We say C is a **smooth** curve if there are no singular points in $C(K)$ for any field K containing k.

Without proof we mention that for $k = \mathbb{Q}$ or $k = \mathbb{R}$ it is sufficient to check that $C(\mathbb{C})$ has no singular points. Luckily, for the curves we will be interested in even $C(k)$ will be enough.

4.3 Bézout's theorem

Theorem 4.7. *Let C be a projective curve of degree d defined over k and let L be a line defined over k. Suppose there is a point in $L(k)$ that does not belong to $C(k)$. Then*

$$\sum_{P\in C\cap L} I_P(C,L) \leqslant d$$

If k is algebraically closed the above sum is equal to d.

Example. Let $C : F = X^3 + X^2Z + XZ^2 + Z^3 - Y^2Z = 0$ be a cubic curve defined over $\mathbb{Q}$ and let L be given by the parametrisation $\ell(s : t) = (2s + 2t : s + 3t := s + t)$. If we plug it into the equation of C, we obtain, after some simplifications

$$0 = F\bigl(\ell(s : t)\bigr) = 6s^3 + 26s^2t + 26st^2 + 6t^3 = 2(3s + t)(s + t)(s + 3t).$$

Hence there are three points of intersection $\ell(1 : -3) = (1 : 2 : 1)$, $\ell(1 : -1) = (0 : 1 : 1)$ and $\ell(-3 : 1) = (1 : 0 : 1)$. Each has intersection multiplicity 1; hence we have equality in the statement of the theorem.

If we take any other line L, we always end up with a cubic homogeneous equation in $(s : t)$. Of course this may factor into three linear terms then the sum of the intersection multiplicities will be 3. Or it course split as an irreducible quadratic times a linear factor in which case we have only one point of multiplicity 1. Or it could be an irreducible cubic polynomial in which case our sum is 0. In extreme cases, we the cubic polynomial may split into linear terms but two or even all three terms are equal. Then we would have only two or one point of intersection but when counted with multiplicity as in the theorem, we still end up with an equality in the theorem. ◇

Proof. Pick a parametrisation $\ell(s : t)$ for the line L. Let $F \in k[X, Y, Z]$ be the degree d homogeneous polynomial defining C. Now consider

$$G(s, t) = F\bigl(\ell(s : t)\bigr) \in k[s, t]$$

as a polynomial in two variables s and t. As each monomial in F is of total degree d and the coordinates of $\ell(s : t)$ are linear homogeneous polynomials of degree 1, the polynomial $G(s, t)$ is homogeneous of degree d, say

$$G(s, t) = a_0\, s^d + a_1\, s^{d-1}t + \cdots + a_{d-1}\, st^{d-1} + a_d\, t^d.$$

Note that the hypothesis $L(k) \not\subset C(k)$ assures that there is a choice of $(s : t) \in \mathbb{P}^1(k)$ for which $G(s : t) \neq 0$, implying that G is a non-zero polynomial. Now we use Lemma 4.2. Let $\bigl\{(s_0 : t_0), (s_1 : t_1), \dots, (s_e, t_e)\bigr\}$ be the points $(s : t)$ in $\mathbb{P}^1(k)$ such that $G(s,t) = 0$. Then

$$G(s,t) = (t_1 s - s_1 t)^{m_1} \cdot (t_2 s - s_2 t)^{m_2} \cdots (t_e s - s_e t)^{m_e} \cdot H(s,t)$$

for some integers m_1, …, m_e and some homogeneous polynomial $H(s,t)$ without zeros in $\mathbb{P}^1(k)$. It is clear that $m_1 + m_2 + \cdots + m_e \leqslant d$ just by looking at the degree on each side. If k is algebraically closed, then $H(s,t) = 1$ and so $m_1 + m_2 + \cdots + m_e = d$.

Write $P_i = \ell(s_i : t_i)$. We now show that $m_i = I_{P_i}(C, L)$. If $s_i = 0$, this is true directly from the definition otherwise we have to change parametrisation. For this let $\tilde{\ell}(s : t) = \ell(s_i\, s : t_i\, s + t)$. If $s_i \neq 0$, this is again a parametrisation of L, but now $\tilde{\ell}(1 : 0) = \ell(s_i : t_i) = P_i$. To find $I_{P_i}(C, L)$ we want to look at

$$\tilde{G}(s,t) = F\bigl(\tilde{\ell}(s : t)\bigr) = F\bigl(\ell(s_i\, s : t_i\, s + t)\bigr) = G(s_i\, s : t_i\, s + t).$$

Write $G(s,t) = (t_i s - s_i t)_i^m \cdot H_i(s,t)$ with $H_i(s_i, t_i) \neq 0$. We get

$$\tilde{G}(s,t) = (t_i s_i s - s_i(t_i s + t))^{m_i} H_i(s_i\, s : t_i\, s + t) = t^{m_i} \cdot s_i^{m_i}\, H_i(s_i\, s : t_i\, s + t)$$

which shows that $I_{P_i}(C, L) = m_i$ because t can not be a factor of $H_i(s_i\, s : t_i\, s + t)$. □

Theorem 4.8. *Let C be a conic defined over k. Then C is smooth if and only if it is non-degenerate.*

Proof. We will proof that C, given by the general equation (4.2), having a singular point $P \in C(K)$ is equivalent to C being degenerate with the two lines meeting at P.

First if C is degenerate and it consists of two lines, then the point where they meet is a singular point. Conversely, let P be a singular point on C. We wish to pick Q to be a point on C different from P. Here is where we may need a field larger than k: Assume first $a_0 \neq 0$. Pick any Y_1 different from the Y-coordinate of P. Then for some field K the equation $F(X, Y_1, 1) = 0$, which is a quadratic equation in X, will have a solution; for instance in an algebraically closed field K containing k this is true. If $a_0 = 0$, we can do this fixing X or Z instead.

Take L to be the line through P and Q. By hypothesis $I_Q(C, L) > 1$ and $I_P(C, L) > 2$ as P is singular. However this contradicts that the sum $\sum_{R \in C \cap L} I_R(C, L) \leqslant 2$. Therefore $L \subset C$. Using Lemma 4.3, we conclude that L is a component of C and hence C is degenerate. □

End of proof of Theorem 4.4. In the proof of Theorem 4.4, we had reached the point where we could assume that all three coefficients in the equation (4.4) are zero. Ah, but now, we recognise equation (4.4) as the equation of the tangent at the point P to the curve C. By assuming that all coefficients are zero, we get that P is a singular point by Theorem 4.6. Hence by the previous theorem, C is degenerate, but that we excluded in the hypothesis of Theorem 4.4. □

The generalisation of the above Theorem 4.7 is called Bézout's Theorem. It states that two curves C_1 and C_2 of degree d_1 and d_2 intersect in at most $d_1 \cdot d_2$ points. More precisely if the field k is algebraically closed and the intersection is counted with multiplicity, then there are exactly $d_1 \cdot d_2$ intersection points. An example is illustrated in Figure 14. The general formula looks like

$$\sum_{P \in C_1 \cap C_2} I_P(C_1, C_2) \leqslant d_1 \cdot d_2$$

for an appropriate definition of the intersection multiplicity $I_P(C_1, C_2)$. Note that this is wrong if we had not taken projective curves (because we would miss out the intersection points at infinity) and it also needs that the two curve do not have a common component (for instance both containing a line as a factor of the defining homogeneous polynomial would not work). The proof of the general statement uses quite a bit of commutative algebra.

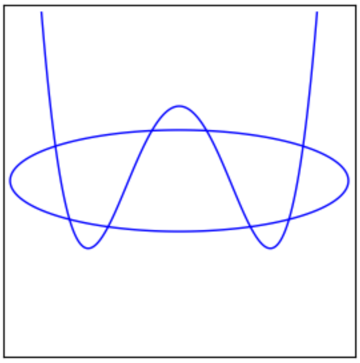


Figure 14: A curve of degree 4 intersects a curve of degree 2 in 8 points

5 Elliptic Curves

Recall that a cubic curve is a projective curve of degree 3. A general cubic curve can be given by a homogeneous polynomial

$$\begin{multline} b_0\,X^3 + b_1\,X^2Y + b_2\,X^2Z + b_3\,XY^2 + b_4\,XYZ + \\ + b_5\,XZ^2 + b_6\,Y^3 + b_7\,Y^2Z + b_8\,YZ^2 + b_9\,Z^3 = 0 \quad (5.1) \end{multline}$$

with constants $b_0, \ldots, b_9$ in k. Again we assume that at least one coefficient is non-zero.

Definition. An **elliptic curve** E defined over k is a non-singular cubic curve E together with a fixed point $O \in E(k)$.

We see already in this definition that an elliptic curve is really a pair (E, O), but we will very often only write E.

Let us start with a very, very important example: Take the equation

$$E: \qquad Y^2Z = X^3 + AXZ^2 + BZ^3 \qquad (5.2)$$

with constants A and B in k. It is more common to see this equation written as an affine curve

$$y^2 = x^3 + Ax + B. \qquad (5.3)$$

Such an equation is called a **Weierstrass equation**. In this case we have a convenient choice of the point O, namely $(0 : 1 : 0)$, which happens to be the unique point at infinity on E. Two pictures of such curves are given in Figure 15.

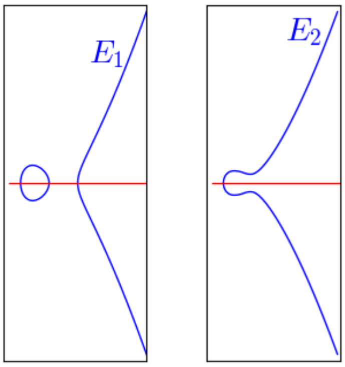


Figure 15: Weierstrass equations can have one or two "pieces" over $\mathbb{R}$

What is the condition that E is non-singular? Well, the conditions for a point $(X : Y : Z)$ to be singular is given in Theorem 4.6 using the partial derivatives:

$$\begin{aligned} 0 &= 3X^2 + AZ^2 \\ 2YZ &= 0 \\ Y^2 &= 2AXZ + 3BZ^2 \end{aligned}$$

Assume that the characteristic of k is not 2, then the second equation says $Y = 0$ or $Z = 0$.

First, if $Z=0$, then we have the unique point $O=(0:1:0)$. But this will not satisfy the third equation above.

So we can take $Y=0$. Multiply the Weierstrass equation (5.2) by A and use the first equation above in the form $AZ^2=-3X^2$. This yields

$$0=AX^3+AX(-3X^2)+BZ(-3X^2)=-(2AX+3BZ)\,X^2.$$

If we had $X=0$, then $Z=0$, too, but $(0:0:0)$ is not a projective point. Therefore $2AX+3BZ=0$. Thus

$$4A^3X^2=A(2AX)^2=A(-3BZ)^2=9B^2AZ^2=9B^2(-3X^2)=-27B^2X^2$$

and using $X\neq 0$ again, we find $4A^3=-27B^2$.

Definition. The **discriminant** of the Weierstrass equation (5.2) is given by $\Delta=-16\cdot(4A^3+27B^2)$.

Lemma 5.1. *Let E be given by the Weierstrass equation (5.2) with A and B in k. Suppose that $\Delta\neq 0$. Then $\bigl(E,(0:1:0)\bigr)$ is an elliptic curve defined over k.*

Proof. If $\Delta\neq 0$ then the characteristic of k is not 2 and therefore $\Delta\neq 0$ is equivalent to $4A^2\neq -27B^2$. We have shown above that in this case E is smooth. □

Example. The curve $C: X^3+Y^3+Z^3=0$ with the point $O=(1:-1:0)$ is an elliptic curve. More generally $C: X^3+Y^3+A\,Z^2=0$ with the point $(1:-1:0)$ is an elliptic curve for all choices of $A\neq 0$ in k. ◇

Example. Consider the cubic curve $C: X^3+7\,Y^3+49\,Z^3=0$ over $k=\mathbb{Q}$. It is smooth. However there is no rational point in $C(\mathbb{Q})$: To prove that assume $(X:Z:Y)\in C(\mathbb{Q})$. Then by multiplying with a common denominator and dividing by any common divisors, we may assume X, Y, Z are three integers and that $\gcd(X,Y,Z)=1$. Considering the equation modulo 7, we find that X is a multiple of 7, say $X=7\,X'$. After dividing by 7, we find $49\,X'^3+Y^3+7\,Z^3=0$. This implies that Y is divisible by 7, say $Y=7\,Y'$. As before, we reach the equation $7\,X'^3+49\,Y'^3+Z^3=0$, which shows that Z is also divisible by 7. However, we divided by the common divisor, hence this is a contradiction. Therefore, we cannot make this curve into an elliptic curve over $\mathbb{Q}$.

Of course $\bigl(C,(\sqrt[3]{7}:-1:0)\bigr)$ is an elliptic curve over $\mathbb{R}$. Also $\bigl(C,(1:1:2)\bigr)$ is an elliptic curve over $\mathbb{F}_5$. ◇

The last example shows that it is not always possible to turn a smooth cubic into an elliptic curve because we could have $C(k) = \varnothing$. It is in fact a very difficult problem to determine, for a give smooth cubic C defined over $\mathbb{Q}$, whether $C(\mathbb{Q}) = \varnothing$ or not. We will return to this at the very end in Section 7.3.

We conclude this section with a formal lemma about the chord and tangent method from Section 1:

Lemma 5.2. *Let C be a smooth cubic defined over k and let P and Q be two distinct points in $C(k)$. Let L be the line through P and Q. Unless L is tangent to C at P or Q, the line L will intersect C in a third point R also defined over k. Furthermore, unless P is an inflection point, the tangent $T_P(C)$ will intersect L in a second point R' also defined over k.*

The proof is already explained in the example right after Bézout's theorem 4.7.

Proof. Parametrise L by $\ell(s:t)$ and say C is given by $F = 0$ for a homogeneous polynomial $F \in k[X,Y,Z]$ of degree 3. Then $G(s,t) = F\bigl(\ell(s:t)\bigr)$ is a homogeneous polynomial of degree 3 unless it is the zero polynomial. The latter only happens when $L \subset C$, which is impossible for a smooth cubic as we saw in an exercise. By assumption we already have two distinct solutions of $G(s,t)$ corresponding to P and Q and they can be of multiplicity strictly higher than 1 as L is not tangent at P or Q. Therefore $G(s,t)$ must factor as

$$G(s,t) = (t_P\, s - s_P\, t) \cdot (t_Q\, s - s_Q\, t) \cdot H(s,t)$$

where $\ell(s_P : t_P) = P$ and $\ell(s_Q : t_Q) = Q$. As the degree is 3, the polynomial $H(s,t) \in k[s,t]$ is linear and it will correspond to a point $R \in L(k)$.

The same reasoning works for $T_P(C)$. This time $G(s,t) = F\bigl(\tilde{\ell}(s:t)\bigr)$ for a parametrisation $\tilde{\ell}$ of $T_P(C)$ will factor as $G(s,t) = (t_P\, s - s_P\, t)^2\, H(s,t)$. Again we get a point $R' \neq P$ in $L(k)$. □

5.1 The group law

Let (E, O) be an elliptic curve defined over k. We will define an operation

$$+ : E(k) \times E(k) \to E(k)$$

which will turn out to be a group law.

Let P and Q be two points in $E(k)$. In a first case, we assume that P, Q and O are three distinct points and suppose they do not lie on one line.

- Draw the line L through P and Q.

- By Lemma 5.2, L meets C in a third point $R \in E(k)$.
- Draw the line L' through R and O.
- Again there is a third point on $L' \cap C$, which we will denote by $P + Q \in E(k)$.

By hypothesis $R \neq O$ and by Lemma 5.2, $P + Q$ has indeed coordinates in k. This is illustrated in Figure 16.

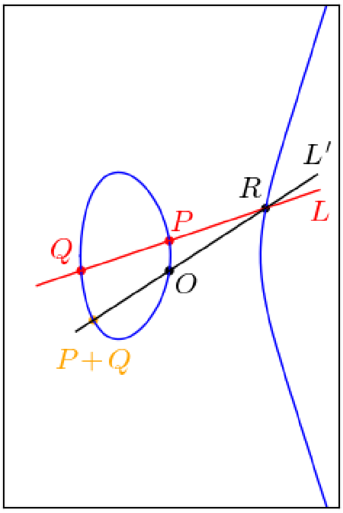


Figure 16: Addition

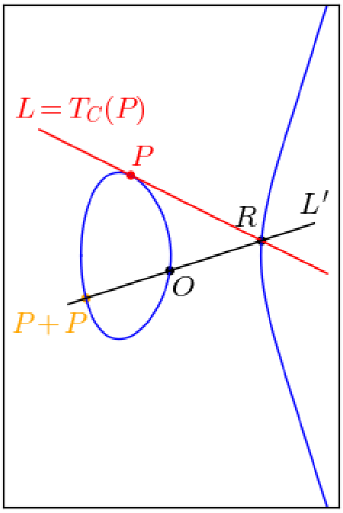


Figure 17: Doubling

We adapt the above to the cases, we excluded. First if P, Q and O lie on one line. Then we take L' to be the tangent $T_O(C)$ and set $P + Q$ to be the second intersection of L' with C guaranteed by Lemma 5.2, unless if O is an inflection point, in which case we take $P + Q$ to be O.

Finally if P and Q are equal but assume they are not inflection points, then we replace in the above L with the tangent $T_P(C)$. This is illustrated in Figure 17. The second point guaranteed by Lemma 5.2 in $T_P(C) \cap C$ will be R and then L' is as before; including in the special cases when $R = O$ discussed above.

Finally when P is an inflection point, then we set $R = P$ and let L' be as before.

All these special cases become very natural when considered as limit cases of the main first case treated. One has just to count multiplicities of intersections between lines and cubics. It is clear that the above does not work when P is a singular point on a non-smooth cubic and Q is non-singular as the line L would not meet in a third point. Also it is clear that O needs to be in $E(k)$ if we want $P+Q$ to be defined over k. This explains why the assumptions on elliptic curves are important.

Theorem 5.3. *The above operation defines an abelian group law on $E(k)$.*

I know at least of three ways to prove this theorem. Firstly one can write down everything in coordinates and check it by hand; this is very, very tedious and hardly illuminating. An excellent proof is using the so-called "Picard group of divisor classes"; it is neat, explains things very well, but uses a lot of algebraic geometry that I would not want to introduce. The third proof is based on classical geometry and that is what we are going to look at here. However the proof will not be given completely as it would take us quite afar from our actual aim.

Incomplete proof. Looking back at Section 2, we see what axioms we have to check. First (G1 Closure): In fact, we discussed this already above when we referred to Lemma 5.2 to assure that in all cases we get exactly one point defined over k. Next (G2 Associativity): Funny enough, this is the really hard part of this theorem and we are postponing this discussion until the end. Let us jump ahead and see that (G5 Abelian) is clear as P and Q play a symmetric role in all cases.

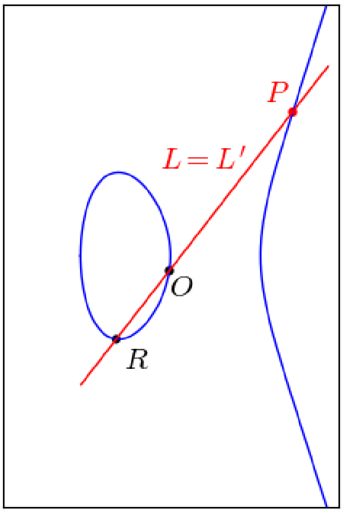


Figure 18: O is the identity element

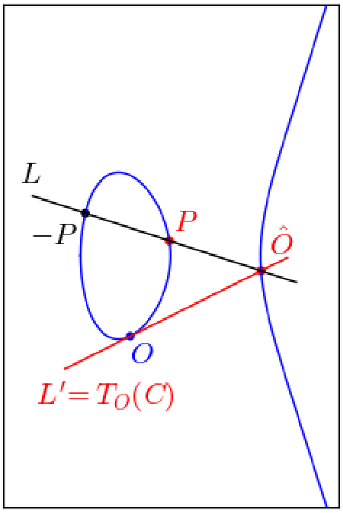


Figure 19: Inverse of P

(G3 Identity): Given the name O, it should be no surprise that $O \in E(k)$ plays the role of the identity element in $E(k)$. See Figure 18. First assume $P \neq O$. Then L is the line through P and O, but so is the line L' and the third point $P + O$ is just P again, even in the special case when L is tangent to C at either P or O. We conclude $P + O = O + P = P$. If $P = O$ then L is going to be the tangent at O. Once again $L' = L$ and we conclude as before that $O + O = O$.

(G4 Inverses): See Figure 19. Start by intersecting the tangent $T_O(C)$ with C. Other than the double intersection at O there is one more point, which we call $\widehat{O}$. If O is an inflection point then $\widehat{O} = O$. Let P be a point in $E(k)$. Let L be the line through P and $\widehat{O}$ and call the third point of intersection $-P$. It may be one of P or $\widehat{O}$ in special cases. To compute $P + (-P)$ we take the line L passing through them and get $R = \widehat{O}$. Then the line L' is $T_O(C)$ and the point $P + (-P)$ identifies with O as we hoped for. This also works in the special cases when $P = O$ or $P = \widehat{O}$.

Finally we come back to (G2 Associativity). We would like to show that $(P + Q) + R = P + (Q + R)$ for all triples P, Q, R of points in $E(k)$. The proof uses following result from classical geometry:

Proposition 5.4 (Chasles' theorem)**.** *Let C, C' and C'' be three distinct projective cubic curves defined over k. Suppose they all pass through a set of eight distinct points $\{P_1, P_2, \ldots, P_8\}$ such that no five of them lie on a line and such that they do not lie all on one conic. Then all three cubics meet in a ninth point P_9.*

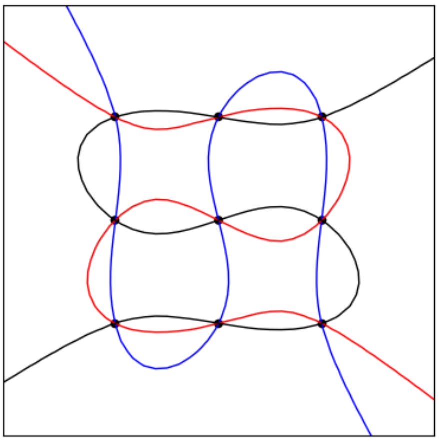


Figure 20: Three cubic meeting in 9 points

This proposition is often called the Cayley-Banarach theorem, see [3] for a detailed history of the theorem and its generalisations. For a complete proof please look at Washington's book [7].

See Figure 20 for an illustration. Here we just sketch the idea of the proof:

A cubic is given by an equation (5.1) involving ten constants $b_0, \dots, b_9$. Multiplying with a constant $\lambda \in k^\times$ does not change the cubic so we can think of the set of all cubics as $\mathbb{P}^9$ a nine-dimensional projective space. To say that C passes through a point $P = (X_0 : Y_0 : Z_0)$ is the same as to say $F(X_0, Y_0, Z_0) = 0$. This imposes a linear condition on the ten coefficients. Therefore saying that C passes through the eight points $P_1, \dots, P_8$ imposes eight linear equations on these ten coefficients. The tricky thing is to show that these equations are linearly independent under our assumptions. But granted they are, we find that the set of cubics passing through the eight points has two degrees of freedom, but up to multiplication by a scalar. In other words they form a projective line in $\mathbb{P}^9$. Say $F = 0$ describes C and $F' = 0$ describes C'. Then the equation $F'' = 0$ for C'' can be written as $F'' = sF + tF'$ for some $(s : t) \in \mathbb{P}^1(k)$. Call P_9 the ninth intersection of C and C', which is unique if we believe in the general version of Bézout's theorem. But then $F''(P_9) = sF(P_9) + tF(P_9) = s \cdot 0 + t \cdot 0 = 0$ shows that P_9 also lies on C''.

Now back to the proof of Theorem 5.3. Let P, Q and R be three points in $E(k)$. We suppose that we are in the least special situation; for instance these points are distinct and not on one line. Also all the lines below are assume not to be tangents to C anywhere, etc. The special cases could all be treated one-by-one with similar arguments. Let L_1 be the line through P and Q and let S_1 be the third point in $C \cap L_1$. Let L_1' be the line through O and S_1 meeting C in the third point $P + Q$. Then L_1'' is the line through $P + Q$ and R, hitting C in a third point U_1.

Let L_2 be the line through Q and R meeting the curve C in the third point S_2. Then L_2' is the line through O and S_2 which contains a third point on C equal to $Q + R$. Finally the line L_2'' goes through $Q + R$ and P and meets C in a further point U_2. Picture? See Figure 21. Note that $(P + Q) + R$ is the point where the line through U_1 and O meets the curve a third time and $P + (Q + R)$ is the point where the line through U_2 and O intersects C. Hence it is enough to show that $U_1 = U_2$.

To apply Chasles' theorem in Proposition 5.4, we need three cubic and eight points. The eight points are O, P, Q, R, $P + Q$, $Q + R$, S_1 and S_2. The first cubic is C. The second cubic is $C' = L_1 \cup L_2' \cup L_1''$, which are precisely the fully drawn lines in Figure 21, and the third is $C'' = L_2 \cup L_1' \cup L_2''$, consisting of the dashed lines in Figure 21. Then the ninth point of intersection of C and C' is

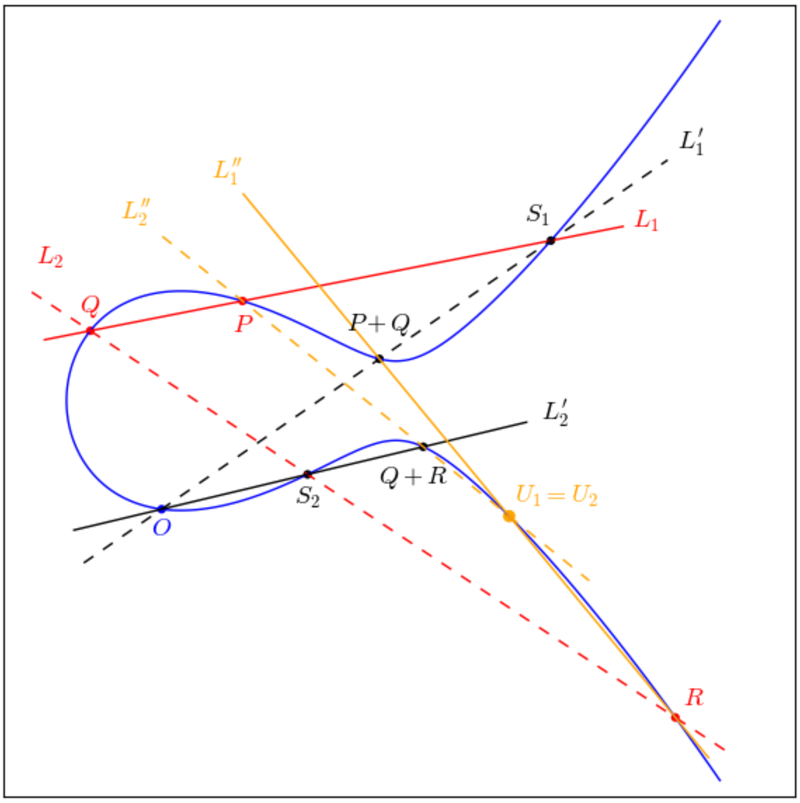


Figure 21: The law is associative

U_1 while U_2 is the ninth point of intersection of C and C''. By the proposition $U_1 = U_2$, proving associativity. Well, at least it concludes our rough overview of what a proof may look like; there are quite a few bits missing. □

5.2 Formula on Weierstrass equations

In this section (for most of the rest of the module) we will assume that our elliptic curve (E, O) is given in a Weierstrass equation

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

with A and B in the field k. We know that this is an elliptic curve with origin $O = (0 : 1 : 0)$ the only point at infinity if and only if $\Delta \neq 0$. Let P and Q be two distinct points in $E(k)$ that are not equal to O. We wish to find the formula for adding P and Q and for $P+P$ in terms of A, B and the coordinates of P and Q.

We have seen that the line at infinity intersects E only in O. It is therefore the tangent to E at O and O is an inflection point. This will simplify things a bit. Also note that all lines through O are of the form $aX + cZ = 0$, which in affine coordinates are simply the vertical lines $x = c$ for a constant $c \in k$.

First, if $O \neq P \in E(k)$, then $-P$ is very easy to describe: Since O is an inflection point, the point $\hat{O}$ is equal to O. Therefore $-P$ and P and O lie on one line, meaning that $-P$ is just the reflection of P on the x-axis. In a simple formula, if $P = (x_P : y_P : 1)$, then $-P = (x_P : -y_P : 1)$. See Figure 22.

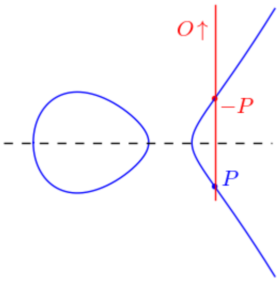


Figure 22: Inverses on a Weierstrass equation

Write $P = (x_P : y_P : 1)$ and $Q = (x_Q : y_Q : 1)$ with x_P, x_Q, y_P, y_Q all in k. The line through P and Q is

$$(y_P - y_Q)\,X + (x_Q - x_P)\,Y + (x_P y_Q - y_P x_Q)\,Z = 0.$$

In affine coordinates we can write this as $y = \lambda\,x + \nu$ with

$$\lambda = \frac{y_P - y_Q}{x_P - x_Q} \qquad \text{and} \qquad \nu = \frac{x_P y_Q - y_P x_Q}{x_P - x_Q}.$$

Inserting into the affine Weierstrass equation yields

$$\begin{aligned}(\lambda\,x + \nu)^2 &= x^3 + A\,x + B\\ 0 &= x^3 - \lambda^2\,x^2 + (A - 2\lambda\nu)\,x + B - \nu^2.\end{aligned}$$

We know that x_P and x_Q are two distinct solutions to this while x_R is the third one. From

$$x^3 - \lambda^2\,x^2 + (A - 2\lambda\nu)\,x + B - \nu^2 = (x - x_P)(x - x_Q)(x - x_R)$$

we get an interesting relation by looking at the degree 2 term: $-\lambda^2 = -x_P - x_Q - x_R$, which means that

$$x_R = \lambda^2 - x_P - x_Q \qquad \text{with } \lambda = \frac{y_P - y_Q}{x_P - x_Q}. \tag{5.4}$$

To get the point R, it is best to use the line equation $y_R = \lambda\,x_R + \nu$. To obtain $P+Q$, we need to look at the line through R and O, but this is simply $x = x_R$. Therefore $P + Q = (x_R : -y_R : 1)$. See Figure 23 for a sketch.

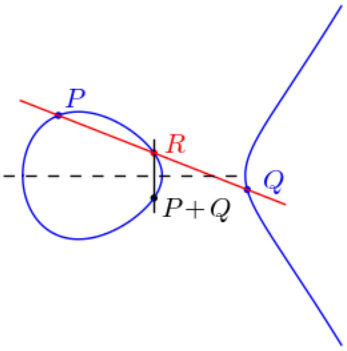


Figure 23: Adding on Weierstrass equations

Example. An example consider the curve $E : y^2 = x^3 - 2\,x + 5$ over $\mathbb{Q}$ and the points $P = (-2 : 1 : 1)$ and $Q = (1 : 2 : 1)$. First we compute $\lambda = (1-2)/((-2)-1) = \frac{1}{3}$. To find ν, it is quickest to use

$$\nu = y_P - \lambda\, x_P = 1 - \frac{1}{3}(-2) = \frac{5}{3}.$$

The first coordinate of R is $x_R = (1/3)^2 - (-2) - 1 = \frac{10}{9}$. It follows that $y_R = \frac{1}{3} \cdot \frac{10}{9} + \frac{5}{3} = \frac{55}{27}$. Therefore $P + Q = \left(\frac{10}{9}, -\frac{55}{27}\right)$.

Let us make a second example to compute $P+(-Q)$, which we will of course write now as $P-Q$. Since $-Q = (1 : -2 : 1)$, the computations look as follows:

$$\lambda = \frac{1-(-2)}{-2-1} = -1 \qquad \nu = 1 - (-1)\cdot(-2) = -1$$
$$x_R = (-1)^2 - (-2) - 1 = 2 \qquad y_R = (-1)\cdot 2 + (-1) = -3$$

Hence $P - Q = (2 : 3 : 1)$. ◇

Oops, we have been careless in deriving the formula (5.4). We divided by $x_P - x_Q$ without checking if it might be zero. All other operations were fine. If $x_P = x_Q$ then either $P = Q$, which we excluded, or $Q = -P$. In the latter case we expect $P + Q = P + (-P) = O$ so it is indeed good if that does not result in a point on the affine plane.

That was adding, now to doubling. As before $P = (x_P : y_P : 1)$. The tangent at P is

$$(3x_P^2 + A)\,X + (-2y_P)\,Y + (-y_P^2 + 2A\,x_P + 3B)\,Z = 0$$

or in affine form $y = \lambda\, x + \nu$ it has

$$\lambda = \frac{3\, x_P^2 + A}{2\, y_P}.$$

Again, we intersect this line with E. As above the quadratic term gives $-\lambda^2 = -x_P - x_P - x_R$. Therefore $x_R = \lambda^2 - 2\, x_P$. The rest is done as before. The Figure 24 illustrates this.

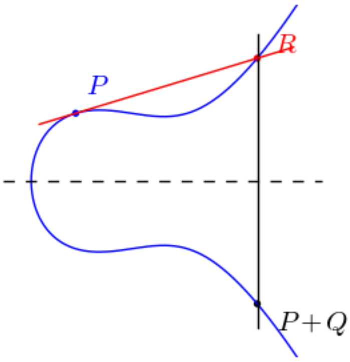


Figure 24: Duplication on Weierstrass equations

Example. We return to the example above and we wish to evaluate $P + P$, which should be denoted by $2\, P$. First, we find the slope of the tangent to be

$$\lambda = \frac{3 \cdot (-2)^2 + (-2)}{2 \cdot 1} = 5$$

and use P to find the constant term of the tangent as

$$\nu = y_p - \lambda x_P = 1 - 5 \cdot (-2) = 11.$$

Then we find the coordinates of R:

$$x_R = 5^2 - 2 \cdot (-2) = 29 \qquad \text{and} \qquad y_R = 5 \cdot 29 + 11 = 156.$$

Therefore $2\, P = (29 : -156 : 1)$ ◇

Again, we lacked to be careful with the division by $2y_P$. Having assumed that $\Delta \neq 0$, it is ok to divide by 2, but what about y_P? If $y_P = 0$, then the tangent is $x = x_P$ which meets the curve a third time at O. It follows that for these point lying on the axis of symmetry, one has $P + P = O$.

The above gives a practical way to compute on an elliptic curve in Weierstrass form, but it could be useful for later if we had an actual general formula. After a bit of simplifying, one finds

$$x_{2P} = \frac{x_P^4 - 2Ax_P^2 - 8Bx_P + A^2}{4y_P^2}. \tag{5.5}$$

Replacing y_P^2 with $x_P^3 + Ax_P + B$ we actually find a formula only involving x_P, A and B.

Example. We end this section with a complete determination of $E(k)$. We take the curve

$$E : y^2 = x^3 - x \tag{5.6}$$

over $k = \mathbb{F}_5$. The discriminant is $\Delta = 4$ so this is a smooth cubic. Look at all possible x value to find all point in $E(\mathbb{F}_5)$:

x	∞	0	1	2	3	4
points	O	$(0,0)$	$(1,0)$	$(2,1)$, $(2,4)$	$(3,2)$, $(3,3)$	$(4,0)$

We count 8 points. Since there are three distinct abelian groups of order 8 we have to do some computations in order to determine the full group structure.

First, the points with the same x-coordinates are inverses to each other. So $-(2,1) = (2,4)$, $-(3,2) = (3,3)$ and $-(0,0) = (0,0)$ and so forth. We find $2\,(0,0) = (0,0) + (0,0) = O$ and $2\,(1,0) = O$ and $2\,(4,0) = O$; these are all points of order 2.

We compute $2\,(2,1)$ using the formula. First $\lambda = (3 \cdot 2^2 + (-1))/(2 \cdot 1) = 11/2 = 11 \cdot 3 = 3$ using the fact that the inverse of 2 modulo 5 is 3. Next, the x-coordinate of $2\,(2,1)$ is $x_R = 3^2 - 2 \cdot 2 = 5 = 0$. Now there is already no choice any more: we must have $2\,(2,1) = (0,0)$. We conclude that $(2,1)$ is an element of order 4.

We suspect now that $P = (2,1)$ and $Q = (1,0)$ are generators of the group. In fact an abstract argument proves this, but let us do it once explicitly. We already know O, Q, P, $2P = (0,0)$ and $3P = -P = (2,4)$. We add P and Q: For λ we have $(1-0)/(2-1) = 1$. Then $x_R = 1^2 - 2 - 1 = 3$. Next $\nu = 1 - 1 \cdot 2 = 4$ and $y_R = 1 \cdot 3 + 4 = 2$. Therefore $P + Q = (3,3)$.

It follows that $3P + Q = -P - Q = -(P+Q) = (3,2)$. The only point left is $(4,0)$ which must be $2P + Q$: Indeed the line through $2P = (0,0)$ and $(1,0)$ is the x-axis which meets E a third time at $(4,0)$.

We conclude that all points in $E(\mathbb{F}_5)$ can uniquely be written as $i\,P + j\,Q$ with $0 \leqslant i < 4$ and $0 \leqslant j < 2$. Or as a formula $E(\mathbb{F}_5) = {}^{\mathbb{Z}}\!/_{4\mathbb{Z}}\,P \times {}^{\mathbb{Z}}\!/_{2\mathbb{Z}}\,Q$. ◇

5.3 Reduction modulo p

We concentrate on the case $k = \mathbb{Q}$. We fix a prime number p. The discussion would be valid for many prime ideals in ring without zero-divisors.

If $P = (X : Y : Z)$ with rational numbers X, Y and Z. Let d be the common denominator. Then $P = (dX : dY : dZ)$ has now integer coordinates. Let e be the greatest common divisor of dX, dY and dZ. Now $P = (d/eX : d/eY : d/eZ)$ is written with integer coordinates which have no common divisor. We call this a **normalised representation** of P.

Example. The normalised representation of $(\frac{3}{2} : 9 : -3)$ is $(1 : 6 : -2)$, which we obtained by scaling with $\frac{2}{3}$. Similarly the point $(-100 : -200 : -300)$ has normalised representation $(-1 : -2 : -3)$; though also $(1 : 2 : 3)$ obtained by scaling with $-\frac{1}{10}$ is also a normalised representation. ◇

We describe the reduction map $\mathbb{P}^2(\mathbb{Q}) \to \mathbb{P}^2(\mathbb{F}_p)$. Let $P = (X : Y : Z)$ be a point in $\mathbb{P}^2(\mathbb{Q})$ given in normalised presentation. We define $\tilde{P} = (\tilde{X} : \tilde{Y} : \tilde{Z})$ with $\tilde{X} = X + p\mathbb{Z} \in \mathbb{F}_p$, $\tilde{Y} = Y + p\mathbb{Z}$ and $\tilde{Z} = Z + p\mathbb{Z}$ the residue classes of the coordinates.

Lemma 5.5. *For any $P \in \mathbb{P}^2(\mathbb{Q})$, defines a well-defined point $\tilde{P}$ in $\mathbb{P}^2(\mathbb{F}_p)$.*

Proof. First we rule out that $\tilde{X} = \tilde{Y} = \tilde{Z} = 0$. If it were so p would divide X, Y, Z, which is impossible as we divided by any common divisor. Therefore $\tilde{P} \in \mathbb{P}^2(\mathbb{F}_p)$.

The normalised representation is almost unique: If $(X : Y : Z)$ is normalised then $(-X : -Y : -Z)$ is also normalised and it is the only other normalised representation of the same point. However both points have give the same point

$$\tilde{P} = (\tilde{X} : \tilde{Y} : \tilde{Z}) = (-\tilde{X} : -\tilde{Y} : -\tilde{Z}) = \big(\widetilde{-X} : \widetilde{-Y} : \widetilde{-Z}\big). \qquad \square$$

The map $P \mapsto \tilde{P}$ is called the reduction map $\mathbb{P}^2(\mathbb{Q}) \to \mathbb{P}^2(\mathbb{F}_p)$. Of course there are reduction maps in all dimensions $\mathbb{P}^n(\mathbb{Q}) \to \mathbb{P}^n(\mathbb{F}_p)$.

Let E be an elliptic curve defined over $\mathbb{Q}$ given by a Weierstrass equation $E : y^2 = x^3 + A\,x + B$. We may suppose that A and B are integers. We will write $\tilde{E}$ for the reduced Weierstrass equation over $\mathbb{F}_p$, that is $\tilde{E} : y^2 = x^3 + \tilde{A}\,x + \tilde{B}$.

Proposition 5.6. *If p does not divide $\Delta \in \mathbb{Z}$ then the reduced curve $(\tilde{E}, \tilde{O})$ is an elliptic curve. Furthermore in this case the reduction map induces a map $\tilde{\ } : E(\mathbb{Q}) \to \tilde{E}(\mathbb{F}_p)$ which is a group homomorphism.*

Proof. It is clear that the discriminant of $\tilde{E}$ is the reduction of the discriminant Δ of E. Hence $\tilde{\Delta} = 0$ if and only if p divides Δ. Showing that $\tilde{E}$ is a smooth cubic over $\mathbb{F}_p$ by Lemma 5.1.

Assume $p \nmid \Delta$. Clearly $\tilde{O} = (0 : 1 : 0) \in \tilde{E}(\mathbb{F}_p)$ is a rational point which makes $\tilde{E}$ into an elliptic curve. Without risk of confusion we will also denote it by O. Write $f(P) = \tilde{P}$ for the restriction of the reduction map to $E(\mathbb{Q})$. The reduced point $\tilde{P}$ belongs to $\tilde{E}(\mathbb{F}_p)$, hence $f : E(\mathbb{Q}) \to \tilde{F}(\mathbb{F}_p)$ is a well-defined map. If $P \in E(\mathbb{Q})$ then $f(-P) = f\bigl((X : -Y : Z)\bigr) = (\tilde{X} : -\tilde{Y} : \tilde{Z}) = -f(P)$. Furthermore, if P, Q and R are three points (not necessarily distinct) forming the intersection of E and the line $L : aX + bY + cZ = 0$. We may clear denominators in L and assume that a, b and c are integers not all divisible by p. Therefore there is a reduced line $\tilde{L} : \tilde{a}X + \tilde{b}Y + \tilde{c}Z = 0$ defined over $\mathbb{F}_p$. Again all three points $\tilde{P}$, $\tilde{Q}$ and $\tilde{R}$ lie on $\tilde{L}$ and on $\tilde{E}$. This proves that, if $P + Q + R = O$, then $f(P) + f(Q) + f(R) = O$. We have shown that f is a group homomorphism. □

If $p \nmid \Delta$ then we say that the Weierstrass equation has **good reduction** otherwise we say it has **bad reduction**. The above proposition has an extension to the case of bad reduction: If one restricts the reduction map to the set of points in $E(\mathbb{Q})$ that do not map to the singular point in $\tilde{E}$ then one obtains a group homomorphism with the group of non-singular points on the singular Weierstrass equation $\tilde{E}$.

Example. We consider the curve

$$E : \qquad y^2 = x^3 - 11\,x - 5$$

The discriminant is 74384. The reduction of E at $p = 5$ is exactly the curve $\tilde{E}$ described in (5.6) in a previous example. We recall that the point $P_1 = (2 : 1 : 1)$ of order 4 and the point $P_2 = (1 : 0 : 1)$ of order 2 generate the group $\tilde{E}(\mathbb{F}_5)$ of order 8. Consider the rational points $Q_1 = (-3 : 1 : 1)$ and $Q_2 = (-2 : 3 : 1)$ in $E(\mathbb{Q})$. Now $\tilde{Q}_1 = (2 : 1 : 1)$ is P_1 while $\tilde{Q}_2 = (3 : 3 : 1) = P_1 + P_2$. It follows that Q_2 cannot be of the form $n\,Q_1$ as this is not true for the images in $\tilde{E}(\mathbb{F}_5)$. Even better: Since the reduction map is surjective onto the group ${}^{\mathbb{Z}}\!/_{4\mathbb{Z}} \times {}^{\mathbb{Z}}\!/_{2\mathbb{Z}}$, we conclude that $E(\mathbb{Q})$ can not be a cyclic group. (In fact, $E(\mathbb{Q})$ is isomorphic to $\mathbb{Z} \times \mathbb{Z}$ generated by Q_1 and Q_2, but that is harder to show.) ◇

5.4 Transforming a cubic into a Weierstrass equation

Let C be any smooth cubic curve defined over k and let $P \in C(k)$.

Proposition 5.7. *Suppose that the characteristic of k is not 2 or 3. For any elliptic curve (C, P) there exists a transformation that brings it to a Weierstrass equation* (5.2)*.*

There are two distinct transformations depending on whether the point P is an inflection point on C or not.

First if P is an inflection point on C. Let $L : aX + bY + cZ = 0$ be the tangent line $T_P(C)$. Pick any other line $L' : a'X + b'Y + c'Z = 0$ passing through P and a third line $L'' : a''X + b''Y + c''Z = 0$ not passing through P. Now make the change of variables: $Z' = aX + bY + cZ$, $X' = a'X + b'Y + c'Z$ and $Y = a''X + b''Y + c''Z$. This transformation is invertible because the three lines do not meet in one point and therefore the matrix with rows (a, b, c), (a', b', c') and (a'', b'', c'') is an invertible matrix. Under this transformation the point P is sent to $(0 : 1 : 0)$ and the line at infinity meets the transformed curve C' only at this point O. This imposes that the cubic equation of C' is of the form

$$b_0 X'^3 + Z' \cdot (b_2 X'^2 + b_4 X'Y' + b_5 X'Z' + b_7 Y'^2 + b_8 Y'Z' + b_9 Z'^2) = 0$$

with $b_0 \neq 0$. This is non-singular at O if $b_7 \neq 0$. We can divide through by b_0. Then we can change the variable Z' by multiplying with $Z'' = b_7/b_0 Z'$. We reach an equation of the form

$$Y^2Z + a_1\, XYZ + a_3\, YZ^2 = X^3 + a_2\, X^2Z + a_4\, XZ^2 + a_6\, Z^6 \tag{5.7}$$

(where we dropped the ' for simplicity). Next, we complete the square on the right hand side, by taking the new Y to be $Y + \frac{1}{2}a_1X + \frac{1}{2}a_3Z$. This allows us to achieve $a_1 = 0$ and $a_3 = 0$ using the assumption that the characteristic is not 2 in order to be allowed to divide by 2. Finally one can complete the cube by replacing X by $X + \frac{1}{3}a_2Z$ to obtain the Weierstrass equation as we presented it.

Example. Consider the cubic $C : X^3 + Y^3 + dZ^3 = 0$ where $d \in k$ is some non-zero constant. As a point we may take $P = (1 : -1 : 0)$. The transformation $X = U + V$ and $Y = U - V$ gives $6UV^2 + 2U^3 + dZ^3 = 0$. Scaling by taking $X' = -6dZ$ and $Y' = 36dV = 18d(X - Y)$ and $Z' = U = (X + Y)/2$ yields finally $Y'^2Z' = X'^3 - 432d^2Z'^3$. This is the Weierstrass equation with $A = 0$ and $B = -432d^2$. ◇

We have to explain the second method of how to do the transformation when P is not an inflection tangent. The method is due to Trygve Nagell, see Chapter 8 in [1]. It is illustrated in Figure 25. Let Q be the third point of intersection of the tangent $T_P(C)$.

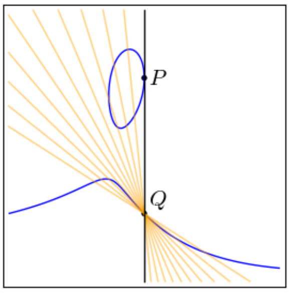


Figure 25: Transforming into Weierstrass form

We may change coordinates such that Q is the affine origin $(0,0)$ and $T_P(C)$ is the y-axis. This means that the curve is given by an equation

$$0 = F_1(X,Y)\,Z^2 + F_2(X,Y)\,Z + F_3(X,Y)$$

with $F_i \in k[X,Y]$ a homogeneous polynomial of degree i. The point $P = (0 : y_P : 1)$ is such that $y = y_P$ is a solution to

$$0 = y\,F_1(0,1) + y^2\,F_2(0,1) + y^3\,F_3(0,1).$$

As it must have a double root one deduces that the discriminant

$$F_2(0,1)^2 - 4\,F_1(0,1)F_3(0,1) = 0 \tag{5.8}$$

is zero. Consider the intersection of C with a general line through Q; in affine form this is $y = tx$. In

$$0 = F_1(1,t)\,x + F_2(1,t)\,x^2 + F_3(1,t)\,x^3$$

we recognise the point Q for $x = 0$. The other two solutions are for $x = (-F_2(1,t) + s)/F_3(1,t)$ with s being a square root of the discriminant:

$$s^2 = F_2(1,t)^2 - 4\,F_1(1,t)\,F_3(1,t).$$

The right hand side of this equation is a cubic polynomial in t because of (5.8). Therefore the transformation with $t = Y/X$ and $s = F_3(1,t)\,X/Z + F_2(1,t)$ transforms C into required form; except for maybe the appearance of a quadratic term on the right hand side. But this can be sorted again by completing the cube.

Example. The computations are often quite involved even for relatively simple examples. We present here the results without details for the cubic curve

$$X^3 + 2\,Y^3 - 3\,Z^3 = 0$$

and the point $P = (1 : 1 : 1)$. The tangent is $X + 2\,Y - 3\,Z = 0$ which meets C again at $Q = (-5 : 4 : 1)$. The change $X' = X + 2\,Y - 3\,Z$ and $Y' = Y - 4\,Z$ and $Z' = Z$ transforms the curve to

$$\begin{aligned}0 = (75\,X - 54\,Y)Z^2 + (-15\,X^2 + 60\,XY - 36\,Y^2)Z+&\\ + (X^3 - 6\,X^2Y + 12\,XY^2 - 6\,Y^3).&\end{aligned}$$

Using the above formula we find the equation

$$s^2 = 72\,t^3 - 216\,t^2 + 216\,t - 75.$$

The final change $s' = 9\,s$ and $t' = 18\,t - 18$ brings this to the Weierstrass equation $s'^2 = t'^3 - 243$. ◇

It is to note that even when the characteristic is 2 or 3, we will always be able to transform our smooth cubic into the form of equation (5.7). This is called the **general Weierstrass equation**. Because of the issue with characteristic 2 and 3 even for equations with integer coefficients it is often better to work with the general Weierstrass equation. For this module, we concentrate mainly on the simpler, also called the short Weierstrass equation (5.2).

We should also add that there is even a more general definition of "elliptic curve". It is a projective smooth curve E defined over k in some $\mathbb{P}^n$, whatever that is precisely, together with a point O and an algebraically defined operation that makes $E(K)$ into a group for all fields K containing k. One can show that a certain invariant called the genus has to be 1 under this condition. The theorem of Riemann-Roch can be used to prove that any such curve can be transformed into a general Weierstrass equation.

In summary, we do not loose any generality if we only consider elliptic curves given by a Weierstrass equation.

6 Torsion points

Let E be an elliptic curve over a field k. We wish to investigate what the group $E(k)$ looks like as an abstract group. We will start by studying the elements of finite order in this group.

Definition. A point $P \in E(k)$ is said to be a **torsion point** if there is an integer $n > 0$ such that

$$n\,P := \underbrace{P + P + P + \cdots + P}_{n \text{ terms}}$$

is equal to O. The smallest such n is called the **order** of P in $E(k)$.

The only point of order 1 is O. If the group $E(k)$ is finite then all points are torsion. This happens for instance when k is a finite field. For fields like $\mathbb{C}$ most points are not torsion; however they are particularly important in the theory of elliptic curves as we might see a glimpse of later.

We assume from now on that it is given by a Weierstrass equation

$$E: \qquad y^2 = x^3 + A\,x + B$$

with A and B in k.

Example. We wish to give an example which is not too simple: The curve $E: y^2 = x^3 - 27\,x + 55350$ over $\mathbb{Q}$ and the point $P = (-21 : 216 : 1)$. The tangent through P has equation $y = 3\,x + 279$ and we find $2\,P = (51, -432)$. The line through P and $2\,P$ has equation $y = -9\,x + 27$ which turns out to be tangent at $2\,P$. Therefore $3\,P = (51, 432) = -2\,P$. Hence $5\,P = O$ shows that P is a point of order 5. We should already be suspicious of how nicely everything stayed within integers and did not involve any denominators. ◇

Definition. If $m > 0$ is an integer then we will write $E(k)[m]$ for the set of points in $E(k)$ such that $m\,P = O$.

We will show in the exercises that $E(k)[m]$ is a subgroup of $E(k)$.

6.1 Points of order two

Lemma 6.1. *The points $(x, y) \in E(k)$ of order 2 are those with $y = 0$.*

Proof. To say that $2\,P = P + P = O$ is the same as to ask that $P = -P$. We know that $-P = (x, -y)$ in a Weierstrass equation. Therefore $P = -P$ is equivalent to $y = 0$. □

If $P = (x_p, 0)$ is a point of order 2 then x_P is a solution to $x^3 + A\,x + B = 0$.

Proposition 6.2. *The subgroup $E(k)[2]$ has either isomorphic to $\{O\}$, ${}^{\mathbb{Z}}/_{2\mathbb{Z}}$ or ${}^{\mathbb{Z}}/_{2\mathbb{Z}} \times {}^{\mathbb{Z}}/_{2\mathbb{Z}}$, depending on whether $0 = x^3 + A\,x + B$ has 0, 1 or 3 solutions in k.*

Proof. If a point is in $E(k)[2]$ then its order is either 2 or 1. From the previous lemma it is clear that we can count the elements of order 2 by counting the number of solutions to $0 = x^3 + A\,x + B$. This tells us the size of $E(k)[2]$. If there are no points of order 2 then $E(k)[2] = \{O\}$. If there is exactly one point of order 2 then $E(k)[2] \cong {}^{\mathbb{Z}}/_{2\mathbb{Z}}$. Finally if there are three then the group is either ${}^{\mathbb{Z}}/_{2\mathbb{Z}} \times {}^{\mathbb{Z}}/_{2\mathbb{Z}}$ or ${}^{\mathbb{Z}}/_{4\mathbb{Z}}$. However all elements have order 2 so it must be $E(k)[2] = {}^{\mathbb{Z}}/_{2\mathbb{Z}} \times {}^{\mathbb{Z}}/_{2\mathbb{Z}}$. □

Let E be an elliptic curve with three points of order 2 defined over k. Then we can write

$$y^2 = (x - e_1)(x - e_2)(x - e_3)$$

for some $e_i \in k$. Then $E(k)[2] = \bigl\{O, (e_1, 0), (e_2, 0), (e_3, 0)\bigr\}$. It is not hard to see that $(e_1, 0) + (e_2, 0) = (e_3, 0)$. The discriminant of E in this form is

$$\Delta = 16(e_1 - e_2)^2(e_2 - e_3)^2(e_3 - e_1)^2. \tag{6.1}$$

In particular the curve is smooth if and only if e_1, e_2, e_3 are all distinct.

Example. Take $k = \mathbb{R}$. Any real cubic polynomial has at least one root. So there is always a point of order 2. If it has three roots e_1, e_2, e_3 in $\mathbb{R}$ then the discriminant in (6.1) is positive. Instead, if there is only one real root e_1 then the other two complex roots are complex conjugates e_2 and $e_3 = \overline{e_2}$. In that case the discriminant

$$\begin{aligned}\Delta &= 16 \cdot \bigl((e_1 - e_2)(e_1 - \overline{e_2})\bigr)^2 \cdot (e_2 - \overline{e_2})^2 \\ &= 16 \cdot \bigl((e_1 - e_2)\overline{(e_1 - e_2)}\bigr)^2 \cdot (2\,i\,\mathrm{Im}(e_2))^2 \\ &= -64 \cdot |e_1 - e_2|^4 \cdot \mathrm{Im}(e_2)^2\end{aligned}$$

is a negative real number.

In summary, if $\Delta > 0$ then $E(\mathbb{R})[2] = {}^{\mathbb{Z}}/_{2\mathbb{Z}} \times {}^{\mathbb{Z}}/_{2\mathbb{Z}}$. If $\Delta < 0$, then $E(\mathbb{R})[2] = {}^{\mathbb{Z}}/_{2\mathbb{Z}}$. This corresponds to the two different real pictures shown in Figure 15. ⋄

6.2 Torsion points have integer coordinates

We now start our work to find a way to determine all torsion points in $E(\mathbb{Q})$ for an elliptic curve over $\mathbb{Q}$. Our first aim is to show that any torsion point $O \neq P = (x : y : 1)$ has integer coordinates x and y. In an exercise, we have shown that this is equivalent to proving that P never reduces to the identity modulo any prime p.

Let us fix a prime p. The exercise referred to above also showed that the normalised representation of a point $P \in E(\mathbb{Q})$ is of the form $(a\,e : b : e^3)$ with $\gcd(a, e) = \gcd(b, e)$. For $n > 0$, define the following subset

$$E_n(\mathbb{Q}) = \Bigl\{ P = (a\,e : b : e^3) \in E(\mathbb{Q}) \Bigm| p^n \mid e \Bigr\}.$$

The first set $E_1(\mathbb{Q})$ of these set is precisely the set of points P that reduce to O modulo p. Since the reduction is a group homomorphism in case of good reduction, $E_1(\mathbb{Q})$ is a subgroup of $E(\mathbb{Q})$ as it is the kernel of this homomorphism. Our aim is to show that all $E_n(\mathbb{Q})$ are subgroups and that none of them contains a torsion point other than O. Note already that these sets form a chain

$$E(\mathbb{Q}) \supset E_1(\mathbb{Q}) \supset E_2(\mathbb{Q}) \supset \cdots$$

and that the only point which belongs to all of them is O. We abbreviate the notation and write simply E_n for $E_n(\mathbb{Q})$.

We introduce a new variable $t = -x/y = -X/Y$. We see that $P \in E_n$ has $p^n \mid t(P) = -x_P/y_P = ae/b$.

Lemma 6.3. *If P, Q, R are points in E_n satisfying $P + Q + R = O$, then $t_P + t_Q + t_R \equiv 0 \pmod{p^{5n}}$.*

Proof. We introduce a second variable $s = 1/y = Z/Y$. Conversely we have $x = -t/s$ and $y = 1/s$. Substituting this into the Weierstrass equation yields

$$\frac{1}{s^2} = \Bigl(-\frac{t}{s}\Bigr)^3 + A\Bigl(-\frac{t}{s}\Bigr) + B.$$

Multiplying with s^3 and rearranging gives

$$t^3 + A\,s^2 t - B\,s^3 + s = 0. \tag{6.2}$$

This is a new affine equation of E. In these coordinates the point O is at the origin $(s, t) = (0, 0)$.

Let us write down how to add points in these coordinates. The slope of the line $s = \lambda\, t + \nu$ through the points P and Q is still

$$\lambda = \frac{s_P - s_Q}{t_P - t_Q}. \tag{6.3}$$

The intersection of the line and E in the equation (6.2) is obtained from

$$0 = t^3 + A\,(t\lambda + \nu)^2 t - B(t\lambda + \nu)^3 + (t\lambda + \nu) = c(t - t_P)(t - t_Q)(t - t_R).$$

As before we look at the quadratic term, but we also need to look at the leading term. We get

$$\begin{aligned} 1 + A\lambda^2 - B\lambda^3 &= c; \\ 2A\lambda\nu - 3B\lambda^2\nu &= c(-t_P - t_Q - t_R). \end{aligned}$$

We have obtained an expression for the quantity that we are after:

$$t_P + t_Q + t_R = \frac{(3B\lambda - 2A)\lambda\nu}{1 + A\lambda^2 - 3B\lambda^3}. \tag{6.4}$$

We will now proceed to prove that λ is divisible by p^{2n} and ν is divisible by p^{3n}, by which we mean that the numerator of these rational numbers are divisible by the corresponding power of p. It will then follow that the numerator of the right hand side of equation (6.4) is divisible by p^{5n} The denominator is not divisible by p since it is congruent to 1 modulo p^{4n}.

As P and Q are in E_n, we know that p^n divides t_P and t_Q. Also $s = e^3/b$ shows that p^{3n} divides s_P and s_Q. From equation (6.3), we see that we have a good chance that p^{2n} divides λ. However it could be that $t_P - t_Q$ is divisible by a higher power than p^n in case they happen to be congruent modulo p^{n+1}. We'd better find another formula for λ.

The idea is to take the difference between the equation (6.2) at the point P and Q.

$$0 = (t_P^3 - t_Q^3) + A(s_P^2 t_P - s_Q^2 t_Q) - B(s_P^3 - s_Q^3) + (s_P - s_Q).$$

This can be rearranged to

$$\begin{aligned} 0 = (t_P^2 + t_P t_Q + t_Q^2)(t_P - t_Q) + &\\ + A\left(s_Q^2(t_P - t_Q) + t_P(s_P + s_Q)(s_P - s_Q)\right) &\\ - B(s_P^2 + s_P s_Q + s_Q^2)(s_P - s_Q) + (s_P - s_Q) & \end{aligned}$$

and then separating $t_P - t_Q$ and $s_P - s_Q$ one obtains

$$(t_P^2 + t_P t_Q + t_Q^2 + A s_Q)(t_P - t_Q) = \\ \bigl(-A t_P(s_P + s_Q) + B(s_P^2 + s_P s_Q + s_Q^2) + 1\bigr)(s_P - s_Q)$$

which finally gets us to

$$\lambda = \frac{s_P - s_Q}{t_P - t_Q} = \frac{t_P^2 + t_P t_Q + t_Q^2 + A s_Q}{1 + B(s_P^2 + s_P s_Q + s_Q^2) - A t_P(s_P + s_Q)}. \tag{6.5}$$

In this form it is now clear that p^{2n} divides λ as the denominator on the right hand side is congruent to 1 modulo p and the numerator is divisible by p^{2n}.

Finally $\nu = s_P - \lambda t_P$ is now divisible by p^{3n} as so is s_P and $\lambda \cdot t_P$.

This deals with the case when $P \neq Q$. We should now redo everything for the tangent at P in case $P = Q$. However, it turns out that we are allowed to set $t_P = t_Q$ and $s_P = s_Q$ in the right hand side of equation (6.5) and we get the correct slope of the tangent, namely

$$\lambda = \frac{3t_P^2 + A s_P^2}{1 + 2A t_P s_P - 3B s_P^2}$$

and now the argument is the same as before. □

Corollary 6.4. *For any $n \geqslant 1$, the set E_n is a subgroup of $E(\mathbb{Q})$. The map $t : E_n \to \mathbb{Z}/p^{5n}\mathbb{Z}$ sending P to $t_P + p^{5n}\mathbb{Z}$ is a group homomorphism.*

Proof. It is clear that $O \in E_n$ and that E_n is closed under taking inverses as $t_{-P} = -t_P$. Let P and Q belong to E_n. Using the notation from the previous proof $s_{P+Q} = -s_R = -\lambda\, t_{P+Q} - \nu$. Now p^{3n} divides ν and $\lambda\, t_{P+Q} \equiv \lambda(t_P + t_Q) \pmod{p^{5n}}$. If $P + Q = (ae : b : e^3)$ then $s_{P+Q} = e^3/b$. This shows that p^n divides e and therefore $P + Q \in E_n$.

Now that E_n is a group to show that t is a group homomorphism it is enough to show that $t(-P) = -t(P)$ and $t(P) + t(Q) + t(R) = 0$. This is now clear from the previous lemma. □

Proposition 6.5. *The only torsion point in the subgroup E_1 is O. In particular, for any prime of good reduction, the map $E(\mathbb{Q})_{tors} \to \tilde{E}(\mathbb{F}_p)$ is an injective group homomorphism.*

Proof. Let $O \neq P$ be a torsion point and suppose to the contrary that $P \in E_1$. Say p^n is the highest power of p dividing the numerator of t_P. There is an $m > 1$ such that $mP = O$.

By the above lemma $0 = t_{mP} \equiv m \cdot t_P \pmod{p^{5n}}$. By our choice of n, this means that p^{4n} divides m. Let $m' = m/p$ and set $Q = m' P$. The point Q is now a point of order p. Let p^l be the highest power of p dividing the numerator of t_Q. Then $0 = t_{pQ} \equiv p\, t_Q \pmod{p^{5l}}$. However by our choice of l the highest power dividing $p \cdot t_Q$ is p^{l+1}, which is the contradiction we were looking for. □

Example. The fact that the reduction map is injective on torsion points can often be used effectively to determine them. As a first example take the curve $E : y^2 = x^3 + 2\,x + 5$. The discriminant $\Delta = -11312$ is not divisible by 3 so the reduction at 3 is good. Checking all x in $\mathbb{F}_3$, we find that for this curve $\tilde{E}(\mathbb{F}_3) = \{O\}$ is the trivial group. By the above proposition $E(\mathbb{Q})_{\text{tors}}$ is also trivial. ◇

Example. As a second example, we look at the curve $E : y^2 = x^3 + 4\,x + 2$. The discriminant $\Delta = -5824$ is not divisible by 3. We find all points modulo 3:

$$\tilde{E}(\mathbb{F}_3) = \Big\{O,\ (1:1:1),\ (1:2:1),\ (2:0:1)\Big\}.$$

This time we can only conclude that $E(\mathbb{Q})_{\text{tors}}$ has either 1, 2 or 4 elements. Since Δ is not divisible by 5 we can also look at the reduction modulo 5.

$$\tilde{E}(\mathbb{F}_5) = \Big\{O,\ (3:1:1),\ (3:4:1)\Big\}.$$

This implies that the number of torsion point in $E(\mathbb{Q})$ must also be a divisor of 3. Hence $E(\mathbb{Q})_{\text{tors}}$ is trivial again.

Instead of looking at the reduction modulo 5, we could also argue as follows. If there were non-trivial torsion points, then there would be an point of order 2. But a point P of order 2 has $y_P = 0$ and so $x_P^3 + 4\,x_P + 2 = 0$. But the only real solution of that is about -0.47347; and that is not an integer. Therefore there is no point of order 2 and hence no non-trivial torsion point in $E(\mathbb{Q})$. ◇

Corollary 6.6. *Any non-zero torsion point in $E(\mathbb{Q})$ is an integral point.*

We add a remark that the above idea of proof even works for the general Weierstrass equation (5.7) if the coefficient a_1 is zero. Otherwise there can be denominators. For instance the curve $E : y^2 + x\,y = x^3 - x^2 + 4\,x - 3$ has only two rational points: O and the point $(\frac{3}{4} : -\frac{3}{8} : 1)$ of order 2.

6.3 The theorem of Lutz and Nagell

The following is a theorem first found independently by Élisabeth Lutz and Trygve Nagell in the 1930ies.

Theorem 6.7. *Let E be an elliptic curve defined over $\mathbb{Q}$ by a Weierstrass equation. Let $O \neq P = (x_P : y_P : 1)$ be a torsion point in $E(\mathbb{Q})$. Then x_P and y_P are integers. Furthermore either $y_P = 0$ or $y_P \mid D$ where $D = -\Delta/16 = 4\,A^3 + 27\,B^2$.*

Proof. It was shown in Corollary 6.6 that x_P and y_P are integers. Suppose now that $y_P \neq 0$, which is equivalent to $2\,P \neq O$. Since $2\,P$ is also a torsion point it must also have integer coordinates. Therefore we know that $\lambda^2 = x_{2P} - 2\,x_P$ is also an integer.

We can do the Euclidean algorithm between the polynomial $x^3 + A\,x + B$ and its derivative $3x^2 + A$ to find the equality

$$D = 9\,(-2Ax + 3B)\,(x^3 + Ax + B) + (6Ax^2 - 9Bx + 4A^2)(3x^2 + A)$$

or alternatively one can just expand the right hand side to verify it. When evaluating this at P, we find

$$D = 9\,(-2Ax_P + 3B)\,y_P^2 + (6Ax_P^2 - 9Bx_P + 4A^2)\,2\,\lambda\,y_P$$

since $\lambda = (3x_P^2 + A)/(2y_P)$. This is an equality involving only integers and hence y_P divides D. □

Theorem 6.7 leaves only finitely many choices for values of y_P for torsion points P.

Corollary 6.8. *The torsion subgroup $E(\mathbb{Q})_{tors}$ is a finite group.*

On the other hand $E(\mathbb{R})_{\text{tors}}$ is either isomorphic to ${}^{\mathbb{Q}}\!/\!_{\mathbb{Z}}$ or to ${}^{\mathbb{Z}}\!/\!_{2\mathbb{Z}} \times {}^{\mathbb{Q}}\!/\!_{\mathbb{Z}}$. Via the map $\theta \mapsto \exp(2\pi i\theta)$ the group ${}^{\mathbb{Q}}\!/\!_{\mathbb{Z}}$ can be identified with an infinite subgroup of the unit circle in $\mathbb{C}$. For $k = \mathbb{C}$ we have always $E(\mathbb{C})_{\text{tors}} = {}^{\mathbb{Q}}\!/\!_{\mathbb{Z}} \times {}^{\mathbb{Q}}\!/\!_{\mathbb{Z}}$ and that is the biggest a torsion subgroup on an elliptic curve can ever be. If you know what p-adic numbers $\mathbb{Q}_p$ are then you might detect that the above theorem is actually proving even that $E(\mathbb{Q}_p)_{\text{tors}}$ is finite.

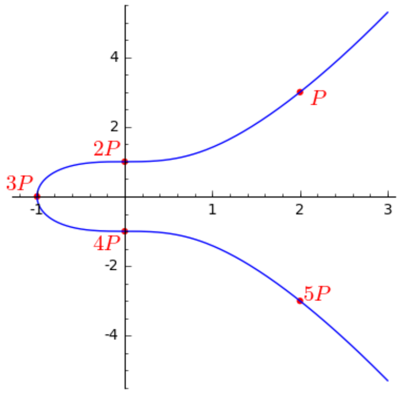


Figure 26: A curve with six torsion points

Example. As a further example, we take the curve $E : y^2 = x^3 + 1$. We have $D = 27$ and so the only possibilities for the y-coordinate of a torsion point are $\{0, \pm 1, \pm 3, \pm 9, \pm 27\}$. We compute the corresponding values of $x_P^3 = y_P^2 - 1$. They are $x_P^3 \in \{-1, 0, 8, 80, 728\}$. But only -1, 0 and 8 are cubes. For each of these points we can check if the point is indeed a torsion point. These gives us the complete list of torsion points

$$E(\mathbb{Q})_{\text{tors}} = \Bigl\{O,\ (-1,0),\ (0,1),\ (0,-1),\ (2,3),\ (2,-3)\Bigr\}.$$

As there is only one abelian group of order 6 this is isomorphic to $\mathbb{Z}/_{6\mathbb{Z}}$. The point $P = (2 : 3 : 1)$ is a generator as one can see quickly from drawing a few lines on the curve as in Figure 26. ◇

6.4 Mazur's theorem

We saw that it is easy to determine when an elliptic curve E defined over $\mathbb{Q}$ has a point of order 2. In principle it is easy to check this for any given order d. There is a polynomial ψ_d of degree $(d^2-1)/2$ whose roots correspond to the x-coordinates of all points of odd order d. These are called division polynomials. For instance for $d = 3$ and $d = 5$, they are

$$\begin{aligned}\psi_3 &= 3x^4 + 6\,A\,x^2 + 12\,B\,x - A^3;\\ \psi_5 &= 5x^{12} + 62A\,x^{10} + 380B\,x^9 - 105A^2\,x^8 + 240AB\,x^7 +\\ &\quad + (-300A^3 - 240B^2)x^6 - 696A^2B\,x^5 + (-125A^4 - 1920AB^2)\,x^4 +\\ &\quad + (-80A^3B - 1600B^3)\,x^3 + (-50A^5 - 240A^2B^2)\,x^2 +\\ &\quad + (-100A^4B - 640AB^3)\,x + A^6 - 32A^3B^2 - 256B^4.\end{aligned}$$

Luckily for $k = \mathbb{Q}$ one never needs very large d. In fact there are no points of order d in $E(\mathbb{Q})$ when $d > 12$. This statement is actually really difficult to prove. The conjecturally complete list of possible groups that can occur as $E(\mathbb{Q})_{\text{tors}}$ was found by Levi in 1908 and rediscovered by Ogg in the 1960ies. It was Barry Mazur in the 1970ies that found a proof of the following theorem using so called modular curves.

Theorem 6.9. *Let E be an elliptic curve defined over $\mathbb{Q}$. Then $E(\mathbb{Q})_{tors}$ will be one of the following fifteen possible abelian groups:*

$$\begin{aligned}\mathbb{Z}/_{d\mathbb{Z}} &\qquad \textit{for } 1 \leqslant d \leqslant 10 \textit{ or } d = 12 \textit{ or}\\ \mathbb{Z}/_{2\mathbb{Z}} \times \mathbb{Z}/_{2d\mathbb{Z}} &\qquad \textit{for } 1 \leqslant d \leqslant 4.\end{aligned}$$

7 Mordell's theorem

The main aim of this section is to prove, under some restriction, that a finite number of initial points in $E(\mathbb{Q})$ are enough to construct all points in $E(\mathbb{Q})$ using chords and tangents. Using the group law, we can state differently that we want to show that $E(\mathbb{Q})$ is a finitely generated abelian group. Note that it is crucial that we work over $\mathbb{Q}$ here; the group $E(\mathbb{R})$ is never finitely generated.

7.1 The weak theorem

Let E be an elliptic curve defined over $\mathbb{Q}$. In this section, we will assume that $E(\mathbb{Q})[2]$ contains 4 points; or equivalently that

$$x^3 + A\,x + B = (x - e_1)(x - e_2)(x - e_3)$$

for some e_1, e_2, $e_3 \in \mathbb{Q}$. We may assume that A and B are integers and so $e_i \in \mathbb{Z}$, too. Saying that E is smooth amounts to assuming that e_1, e_2, and e_3 are distinct.

We will now consider an important map. The target of this map is the quotient ${}^{\mathbb{Q}^\times}\!/_{\square}$ of $\mathbb{Q}^\times$ by its subgroup of squares $\square = \{s^2 \mid s \in \mathbb{Q}^\times\}$. In an exercise we show that we can represent each element uniquely as $u \cdot \square$ where u is a is a square-free integer u.

We define the maps κ_1, κ_2, κ_3 as follows:

$$\begin{aligned} \kappa_i \colon E(\mathbb{Q}) &\to {}^{\mathbb{Q}^\times}\!/_{\square} && (7.1)\\ P = (x,y) &\mapsto (x - e_i)\cdot\square \qquad \text{if } x \neq e_i \text{ and } P \neq O && (7.2)\end{aligned}$$

The two additional values are defined by

$$\kappa_i(O) = 1 \quad \text{and} \quad \kappa_i\bigl((e_i,0)\bigr) = (e_i - e_j)(e_i - e_k)\cdot\square \qquad (7.3)$$

where $\{i,j,k\} = \{1,2,3\}$. Putting the three maps together, we have a big map

$$\begin{aligned} \kappa \colon E(\mathbb{Q}) &\to {}^{\mathbb{Q}^\times}\!/_{\square} \times {}^{\mathbb{Q}^\times}\!/_{\square} \times {}^{\mathbb{Q}^\times}\!/_{\square} && (7.4)\\ P &\mapsto \bigl(\kappa_1(P), \kappa_2(P), \kappa_3(P)\bigr). && (7.5)\end{aligned}$$

It is called the **Kummer map**.

Example. Let E be the elliptic curve $y^2 = x^3 - 43\,x + 42 = (x-1)(x-6)(x+7)$. The discriminant is $\Delta = 16\cdot((1-6)(1-(-7))(6-(-7))^2 = 16\cdot 5^2\cdot 8^2\cdot 13^2 = 2^{10}\cdot 5^2\cdot 13^2$. There are the three two torsion points $T_1 = (1,0)$, $T_2 = (6,0)$ and $T_3 = (-7,0)$. With some luck we also spot $P = (-3,12)$, $Q = (11,30)$

and $R = (19, -78)$ as some further points in $E(\mathbb{Q})$. We compute the Kummer map κ evaluated at these points.

For instance

$$\begin{aligned}\kappa(P) &= \bigl((-3-1)\cdot\Box, (-3-6)\cdot\Box, (-3+7)\cdot\Box\bigr)\\ &= \bigl(-4\cdot\Box, -9\cdot\Box, 4\cdot\Box\bigr) = \bigl(-1,-1,1\bigr).\end{aligned}$$

We represented the images as $u\cdot\Box$ with u square-free and we omitted $\cdot\Box$; we will do likewise in the table below.

point	$T_1 = (1,0)$	$T_2 = (6,0)$	$T_3 = (-7,0)$
κ	$(-10,-5,2)$	$(5,65,13)$	$(-2,-13,26)$
point	$P = (-3,12)$	$Q = (11,30)$	$R = (19,-78)$
κ	$(-1,-1,1)$	$(10,5,2)$	$(2,13,26)$

Here is an example with non-integral coordinates:

$$\kappa(-\tfrac{11}{9}, -\tfrac{260}{27}) = (-\tfrac{11}{9}-1, -\tfrac{11}{9}-6, -\tfrac{11}{9}+7) = (-\tfrac{20}{9}, -\tfrac{65}{9}, \tfrac{52}{9}) = (-5,-65,13).$$

It should already look suspicious that the same numbers kept appearing everywhere. ◇

Proposition 7.1. *The map κ_i is a group homomorphism for all $i = 1,2,3$.*

Example. Back to the elliptic curve in the previous example. For instance $Q = T_1 + P$ and indeed $\kappa(T_1)\kappa(P) = (-10,-5,2)(-1,-1,1) = (10,5,2) = \kappa(Q)$. ◇

Proof. In an exercise we showed that $x^{-1} = x$ in the group $\mathbb{Q}^\times/\Box$. Since the definition of κ_i only involves x, one has $\kappa_i(-P) = \kappa_i(P) = \kappa_i(P)^{-1}$. We are left to show that $\kappa_i(P)\,\kappa_i(Q)\,\kappa_i(R) = \Box = 1$ if P and Q and R lie on one line.

First, if one of the three, say P, is equal to O, then $Q = -R$. In this case $\kappa_i(P)\,\kappa_i(Q)\,\kappa_i(R) = 1\cdot\kappa_i(Q)\,\kappa_i(Q)^{-1} = 1$.

Next to the case when all three are distinct from O and from $(e_i, 0)$. Let $y = \lambda\,x + \nu$ be the line through the three points. We can substitute into the equation

$$(x-e_1)(x-e_2)(x-e_3) - (\lambda\,x+\nu)^2 = (x-x_P)(x-x_Q)(x-x_R) \tag{7.6}$$

the value e_i for x and obtain

$$-(\lambda\,e_i+\nu)^2 = (e_i-x_P)(e_i-x_Q)(e_i-x_R) \in -\kappa_i(P)\,\kappa_i(Q)\,\kappa_i(R)$$

which shows that the product $\kappa_i(P)\,\kappa_i(Q)\,\kappa_i(R)$ is the identity element $\square$ in ${}^{\mathbb{Q}^\times}\!/_{\square}$.

Finally, we have to treat the case when one point, say P, is equal to $(e_1, 0)$, but none of the others is O. The cases $(e_2, 0)$ and $(e_3, 0)$ are very similar to this. In this case the line is of the form $y = \lambda(x - e_1)$. Now we take the above equation (7.6), but we divide it by $x - e_1$:

$$(x - e_2)(x - e_3) - \lambda^2\,(x - e_1) = (x - x_Q)(x - x_R).$$

Setting $x = e_1$ yields $\kappa_1(P) = \kappa_1(Q)\,\kappa_1(R)$.

This proves in all cases that κ_i is a group homomorphism. □

Proposition 7.2. *The kernel of κ is $2\,E(\mathbb{Q}) = \bigl\{2\,P \bigm| P \in E(\mathbb{Q})\bigr\}$.*

Example. Let us illustrate this again in the example. For instance $Q' = \bigl(\frac{58}{9}, \frac{154}{27}\bigr)$ is a point on E. Now κ evaluates to $\bigl(\frac{49}{9}, \frac{4}{9}, \frac{121}{9}\bigr) = (1, 1, 1)$. By this proposition it must be a multiple of another point by 2. Indeed it is $2\,(-P)$. ◇

Proof. The first part is to show that $2\,E(\mathbb{Q}) \subset \ker(\kappa)$. If $Q \in E(\mathbb{Q})$, then by the previous proposition $\kappa(2\,Q) = \kappa(Q)^2 = 1$. That was the easy part, the other inclusion follows from the following lemma: The assumption in that lemma is equivalent to $P \in \ker(\kappa)$ and the conclusion is an explicit point P such that $\pm P = 2\,Q$. This implies $P = 2\,Q$ or $P = 2\,(-Q)$ but in any case $P \in 2\,E(\mathbb{Q})$. □

Lemma 7.3. *Let $O \neq P = (x_P, y_P)$. Assume that for each $i \in \{1, 2, 3\}$ there is some $a_i \in \mathbb{Q}^\times$ with $x_P - e_i = a_i^2$. Let $f = u_0 + u_1\,x + u_2\,x^2 \in \mathbb{Q}[x]$ be the polynomial such that $f(e_i) = a_i$. Then $Q = (u_1 : 1 : u_2)$ is a point on $E(\mathbb{Q})$ such that $2\,Q = P$ or $2\,Q = -P$.*

Example. Again we illustrate with a concrete example from before. We saw that $Q' = \bigl(\frac{58}{9}, \frac{154}{27}\bigr)$ is in the kernel of κ. This lemma should give us a way to find Q'' such that $2\,Q'' = \pm Q'$. By the interpolation formula $f = \frac{1}{78}\,(202 - 19\,x - x^2)$ is such that $f(1) = \frac{7}{3}$, $f(6) = \frac{2}{3}$ and $f(-7) = \frac{11}{3}$. Therefore $Q'' = \bigl(-\frac{19}{78} : 1 : -\frac{1}{78}\bigr) = (19 : -78 : 1)$ is such that $2\,Q'' = -Q'$. Maybe we should be surprised that we did not find P when we know that $2P = -Q'$. Well in fact $Q'' = P + T_3$. Therefore $2\,Q'' = 2\,P + 2\,T_3 = 2\,P$. Clearly the lemma just gives one of eight possible points. ◇

Proof. Notice that such a polynomial exists and is unique by the usual interpolation as long as the values of e_1, e_2 and e_3 are distinct. Consider the quartic polynomial $g(x) = x_P - x - f(x)^2$. By construction $g(e_i) = 0$, which implies that $x^3 + A\,x + B$ divides g. This means that the remainder when dividing $f(x)^2$ by $x^3 + A\,x + B$ is equal to $x_P - x$. Since $x^3 \equiv -A\,x - B$ and $x^4 \equiv -A\,x^2 - B\,x$ modulo $x^3 + A\,x + B$ one finds

$$\begin{aligned} x_P - x \equiv &+u_0^2 + 2\,u_0u_1\,x + (u_1^2 + 2u_0u_2)\,x^2 + \\ &+ 2\,u_1u_2(-A\,x - B) + u_2^2\,(-A\,x^2 - B\,x) \pmod{x^3 + Ax + B}. \end{aligned}$$

Since the remainder when dividing is a unique polynomial of degree at most 2, the above is actually an equality, not just a congruence. This results in three equations for the coefficients in $k[x]$:

$$\begin{aligned} x_P &= u_0^2 - 2B\,u_1u_2; \qquad (7.7)\\ -1 &= 2u_0u_1 - 2A\,u_1u_2 - B\,u_2^2;\\ 0 &= u_1^2 + 2u_0u_2 - A\,u_2^2. \qquad (7.8) \end{aligned}$$

Multiply the middle equation with $-u_2$ and the last by u_1 and add them:

$$u_2 = 2Au_1u_2^2 + Bu_2^3 + u_1^3 - Au_1u_2^2$$

or $u_2 = u_1^3 + Au_1u_2^2 + Bu_2^3$ which is precisely the statement that $Q = (u_1 : 1 : u_2)$ lies in $E(\mathbb{Q})$. Write $x_Q = u_1/u_2$ and $y_Q = 1/u_2$ To complete the proof we use (7.8) to write u_0 also in terms of Q:

$$u_0 = \frac{A\,u_2^2 - u_1^2}{2\,u_2} = \frac{A - x_Q^2}{2\,y_Q};$$

and then we substitute the u_i in (7.7) to discover

$$x_P = \Bigl(\frac{A - x_Q^2}{2\,y_Q}\Bigr)^2 - 2B\,\frac{x_Q}{y_Q^2} = \frac{x_Q^4 - 2A\,x_Q^2 - 8B\,x_Q + A^2}{4\,y_Q^2}$$

and this is precisely the duplication formula (5.5). This shows that $2\,Q = \pm P$. □

Example. Here a further example $E : y^2 = x^3 - 351\,x + 1890 = (x + 21)(x - 6)(x - 15)$. The image of the torsion point $T = (15, 0)$ is

$$\kappa(T) = \bigl(15 + 21, 15 - 6, (15 + 21)(15 - 6)\bigr) = \bigl(36, 9, 36 \cdot 9\bigr) = (1, 1, 1).$$

By the above proposition there exists a point Q such that $2\,Q = T$. Since $2T = O$ then $4\,Q = O$.

The previous lemma can be used to find this point of order 4. The polynomial $f(x) = -\frac{40}{9} - \frac{11}{54}\,x - \frac{1}{162}\,x^2$ satisfies $f(-21) = 6$, $f(6) = 3$ and $f(15) = 0$. Therefore $Q = (-\frac{11}{54} : 1 : -\frac{1}{162}) = (33 : -162 : 1)$ is a candidate and indeed $2\,Q = T$. ◇

Proposition 7.4. *Let S be the set containing -1 and all odd primes p such that $p \mid \Delta$. Let H be the finite subgroup of ${}^{\mathbb{Q}^\times}\!/_{\square}$ generated by S. The image of κ_i is contained in H.*

Example. In the example we followed so far, we have $\Delta = 2^{10} \cdot 5^2 \cdot 13^2$. Hence $S = \{-1, 2, 5, 13\}$. Indeed all the images in of κ_i that we have computed so far were products of these four integers. ◇

Proof. To simplify the notation we assume $i = 1$. Let $P \in E(\mathbb{Q})$ and consider $\kappa_1(P) = c \cdot \square$ with a square-free integer c. Let p be a prime that does not divide Δ. In particular, the reduction of E modulo p is an elliptic curve and therefore the three elements $\tilde{e_1}$, $\tilde{e_2}$, $\tilde{e_3}$ in $\mathbb{F}_p$ are distinct.

We treat first the case that $P \neq O$ and $P \neq (e_1, 0)$. Assume p divides the numerator or the denominator of $x_P - e_1$. Write $x_P = a/e^2$ with coprime integers a and e. Since $x_P - e_1 = (a - e_1\,e^2)/e^2$, either $p \mid a - e_1\,e^2$ or $p \mid e$. However if $p \mid e$ then $p \nmid c$ as an even power of p will divide the denominator of $x_P - e_1$ Therefore p divides $a - e_1\,e^2$.

As $e_2 \not\equiv e_1 \pmod{p}$ the expression $a - e_2\,e^2$ is not divisible by p. Similar for $a - e_3\,e^2$. The Weierstrass equation with $y_P = b/e^3$ gives

$$b^2 = (a - e_1\,e^2)(a - e_2\,e^2)(a - e_3\,e^2).$$

Hence p divides b. We see that the left hand side is divisible by an even power of p, therefore the right hand side, too. We conclude that $x_P - e_1$ is divisible by an even power of p. Therefore c is not divisible by p.

Finally if $P = O$ then $c = 1$. If $P = (e_1, 0)$ then $\kappa_1(P) = (e_1 - e_2)(e_1 - e_3) \cdot \square$. Since the e_i are distinct modulo p also this value of κ will not contain p to an odd power.

In all cases, c is a product of some primes p in S and maybe -1. Therefore $\kappa_1(P) \in H$. □

Theorem 7.5. *If $E(\mathbb{Q})[2] \cong {}^{\mathbb{Z}}\!/_{2\mathbb{Z}} \times {}^{\mathbb{Z}}\!/_{2\mathbb{Z}}$, then the group $E(\mathbb{Q})/2E(\mathbb{Q})$ is a finite group.*

Proof. The map $\kappa : E(\mathbb{Q}) \to {}^{\mathbb{Q}^\times}/_{\square} \times {}^{\mathbb{Q}^\times}/_{\square} \times {}^{\mathbb{Q}^\times}/_{\square}$ is a group homomorphism by Proposition 7.1. This map has kernel $2\,E(\mathbb{Q})$ by Proposition 7.2. The image is in the finite group $H \times H \times H$ by Proposition 7.4. By the first isomorphism theorem, we get an isomorphism $E(\mathbb{Q})/2E(\mathbb{Q}) \to \operatorname{im}(\kappa)$ of finite groups. □

This theorem is called the weak Mordell-Weil theorem. Unfortunately it is not enough for proving that $E(\mathbb{Q})$ is finitely generated. For instance $A = \langle \mathbb{Q}, + \rangle$ is an abelian group such that $A/2A$ is finite, even trivial. We need to show that no element P of $E(\mathbb{Q})$ can be "divided by" 2^n for all $n > 1$, meaning that for all $n > 1$ there is a point $Q_n \in E(\mathbb{Q})$ such that $2^n\, Q_n = P$.

Theorem 7.5 also holds without the assumption that there are 3 rational points of order 2. One way to extend the above proof is presented in [4] where the ring $\mathbb{Q}[x]/(x^3 + Ax + B)$ is used. Extending to the case when there is one rational point of order 2 is presented in [6]. The most general proof uses algebraic number fields given in [1] for instance.

7.2 Heights

Definition. We define the **height** of a rational number $r = \frac{a}{b}$ to be the logarithm of the maximum of $|a|$ and $|b|$ when $\frac{a}{b}$ is the reduced fraction of r:

$$h\Big(\frac{a}{b}\Big) = \log \max\big\{|a|, |b|\big\}.$$

One could say that this is a measure of how much ink one needs to write down the reduced fraction of r. Rational numbers with large height are arithmetically more complicated than those with small height.

Lemma 7.6. *Given a bound $B > 0$. There are only finitely many rational numbers r with $h(r) < B$.*

Proof. It says that $|a| < e^B$ and $|b| < e^B$. That leaves only finitely many options for $r = a/b$. □

Example. The list $\big\{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2\big\}$ contains all rational numbers with height less than 1. If $h(r) < 3$ then $|a|$ and $|b|$ must be less than $e^3 \approx 20.09$; this leaves 511 possible r. ◇

For a non-zero point $P \in E(\mathbb{Q})$ on an elliptic curve in Weierstrass form, we define $h(P) = h(x)$. It makes sense to set $h(O) = 0$. Since for each x there are at most two points P, there is a finite number of points in P with a height lower than a given bound.

There are two technical lemmas about how the height changes under addition and duplication.

Lemma 7.7. *For a fixed elliptic curve E in Weierstrass form, there is a constant C such that $h(2\,P) \geqslant 4h(P) - C$ for all $P \in E(\mathbb{Q})$.*

Note C will depend on A and B, but not on P. In fact, the multiplication by 2 roughly multiplies the number of digits of x by 4. In Figure 27 there is an illustration. The parabola appearing above the numbers there suggest that $h(nP)$ is roughly n^2 times a constant.

$$0$$

$$\frac{1}{4}$$

$$72$$

$$-\frac{287}{1296}$$

$$\frac{43992}{82369}$$

$$\frac{26862913}{1493284}$$

$$-\frac{3596697936}{8760772801}$$

$$\frac{7549090222465}{8662944250944}$$

$$\frac{51865013741670864}{6504992707996225}$$

$$-\frac{173161424238594532415}{310515636774481238884}$$

$$\frac{6005923027069067356081464}{4609517672092049172106561}$$

$$\frac{5168009015065791370340979491 53}{116165201153061098261023776144}$$

$$-\frac{3773654236625347557081848561796712 8}{5794167433504947998768876827435276 9}$$

$$\frac{2341967938277653301633872887424665142771 3}{1225880569761762989368984702749518724883 6}$$

$$\frac{264494523477188261711736621823276820476705417 92}{94660948045863857623125096618373029613545504 01}$$

$$-\frac{38936704813849549996845541108410674936880408883714559}{57067021596582095418792572856507183982778272745812224}$$

$$\frac{4591645941205651758743399376678547640439110366472582976891 20}{1601575954031550798105892264973567722587650266521790971820 81}$$

$$\frac{68959182620311388778577311972229580726760149557647155768372919686 41}{36944651848893877713225624066917526092297582124077894864378991481 00}$$

$$-\frac{192213963958496808206022650324690142827789568666593186602979081002044415320}{296918098432425407525773276985596641548336044615314761541661145231217340801}$$

$$\frac{45889113506746061755939334774844873737209801747894547663449687796544554933032988641}{10000421195400838200377728244194903373980117609644984613889776434542245978910953104}$$

Figure 27: The x-coordinates of $n\,P$ with $P = (0 : 1 : 1)$ on $E : y^2 = x^3 + x + 1$ for $1 \leqslant n \leqslant 20$

We are not going to include the proof of this lemma (and the next one). The above is shown in Section III.3 in [6]. The main idea is to look at the duplication formula (5.5) for a point $P = (a/e^2, b/e^3)$, which can be written as a fraction of two integers:

$$x_{2P} = \frac{a^4 - 2A\,a^2e^4 - 8B\,ae^6 + A^2\,e^8}{4(a^3 + A\,a\,e^4 + B\,e^6)e^2}.$$

If there is not too much cancellation in this fraction then the numerator and the denominator will be roughly the forth power of $\max\{|a|, e^2\}$. Taking logarithms, we should expect that $h(2P)$ is roughly $4\,h(P)$.

Lemma 7.8. *Given an elliptic curve E defined over $\mathbb{Q}$ in Weierstrass form and given a fixed point $P_0 \in E(\mathbb{Q})$. Then there exists a constant C_0 such that*

$$h(P + P_0) \leqslant 2\,h(P) + C_0$$

for all points $P \in E(\mathbb{Q})$.

Again, we emphasise that C_0 will depend on A, B and P_0, but not on P. The proof of this lemma is Section III.2 in [6]; it is based on the explicit formula for adding P_0 to P.

These are now all ingredients for concluding the main theorem of this section.

Theorem 7.9 (Mordell's theorem)**.** *Let E be an elliptic curve defined over $\mathbb{Q}$. The group $E(\mathbb{Q})$ is a finitely generated abelian group.*

Proof. By Theorem 7.5 or its generalisation if $E(\mathbb{Q})[2]$ has less than 4 points, the group $E(\mathbb{Q})/2\,E(\mathbb{Q})$ is finite. For each coset modulo $2\,E(\mathbb{Q})$ we pick an element. This is the list $Q_1, Q_2, \ldots, Q_n$. We will add a few more points to these to get a generating set.

Use Lemma 7.8 with the point $P_0 = -Q_i$. There is a constant C_i depending on i such that $h(R - Q_i) \leqslant 2\,h(R) + C_i$ for all $R \in E(\mathbb{Q})$. Take C' to be the largest of the C_i for $1 \leqslant i \leqslant n$ so that $h(R - Q_i) \leqslant 2\,h(R) + C'$ for all i and R. Let C be the constant given in Lemma 7.7. Let $\mathcal{S}$ be the finite list containing $Q_1, Q_2, \ldots, Q_n$ and all points in $E(\mathbb{Q})$ with height less than $C + C'$.

Let P be any point in $E(\mathbb{Q})$. As it belongs to some coset, there is a point P_1 and an integer $1 \leqslant i_1 \leqslant n$ such that $P = 2\,P_1 + Q_{i_1}$. The point P_1 also belongs to some coset and we can do the same again. We get a chain of points $P_1, P_2, \ldots$ in $E(\mathbb{Q})$ and a chain of indices $i_1, i_2, \ldots$ in $\{1, 2, \ldots, n\}$ such that

$$\begin{aligned} P &= 2\,P_1 + Q_{i_1} \\ P_1 &= 2\,P_2 + Q_{i_2} \\ P_2 &= 2\,P_3 + Q_{i_3} \\ \vdots &\quad \vdots \end{aligned}$$

Intuitively, we should believe that the points $P_1, P_2, \dots$ have decreasing height; this is what we wish to justify now. For any $j > 1$, we find

$$4\,h(P_j) \leqslant h(2\,P_j) + C = h(P_{j-1} - Q_{i_j}) + C \leqslant 2\,h(P_{j-1}) + C' + C.$$

As long as $h(P_{j-1}) \geqslant C + C'$, we have

$$h(P_j) \leqslant \tfrac{1}{2}\,h(P_{j-1}) + \tfrac{C+C'}{4} = \tfrac{3}{4}h(P_{j-1}) - \tfrac{1}{4}\bigl(h(P_{j-1}) - C - C'\bigr) \leqslant \tfrac{3}{4}h(P_{j-1}).$$

As the height decreases, there exists an m such that $h(P_m) \leqslant C + C'$. Hence $P_m \in \mathcal{S}$.

We can substitute recursively, starting with $P = 4\,P_2 + 2\,Q_{i_1} + Q_{i_2}$ until we get to

$$P = 2^m\,P_m + 2^{m-1}\,Q_{i_{m-1}} + \cdots + 2\,Q_{i_2} + Q_{i_1}.$$

This shows that we can express any point $P \in E(\mathbb{Q})$ as a combination of points from our finite list $\mathcal{S}$. Therefore $E(\mathbb{Q})$ is finitely generated. □

7.3 Determination of the rank

From Mordell's theorem, we know that $E(\mathbb{Q})$ is a finitely generated abelian group. In G13GTH=MATH3001 you might see the proof that this implies that there exists an integer r such that $E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r$. The integer r is called the **rank** of E. It is the smallest integer such that there are r rational point P_1, P_2, $\dots$, P_r of infinite order that generate $E(\mathbb{Q})$ together with the finite set of torsion points. Since we know how to get all torsion points, we are left to determine r and the point P_i to find information on $E(\mathbb{Q})$.

Since

$$E(\mathbb{Q})/2E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}}/2\,E(\mathbb{Q})_{\text{tors}} \times \bigl({}^{\mathbb{Z}}\!/_{2\mathbb{Z}}\bigr)^r$$

it is therefore enough to look at the weak Mordell-Weil theorem 7.5 to determine r; we do only need heights if we want the points P_i.

Let us assume that $E(\mathbb{Q})[2]$ contains four elements. It follows that the quotient group $E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}}$ has also four elements. Therefore $E(\mathbb{Q})/2E(\mathbb{Q})$ has 2^{r+2} elements.

Proposition 7.10. *Let h be the number of prime divisors of Δ. Then $r \leqslant 2\,h$.*

Proof. The map κ has its image contained in $H \times H \times H$ by Proposition 7.4. However since $y^2 = (x - e_1)(x - e_2)(x - e_3)$, the image lands even in the subgroup of triples (a, b, c) such that $abc = 1$ in H. Therefore the image of κ_1 and κ_2 already determine the image of κ.

Since H is generated by -1 and all the prime divisors of Δ there are 2^{h+1} elements in H. Therefore

$$2^{r+2} = \#E(\mathbb{Q})/2E(\mathbb{Q}) \leqslant \#H \cdot \#H = 2^{2h+2}$$

which gives the above bound. □

Often this bound is much too large. Here is an idea to decrease the bound in practice. Let a, b, c be three square-free integers whose prime divisors divide Δ and such that abc is a square. We would like to determine if there is a point P such that $\kappa(P) = \bigl(a\cdot\square, b\cdot\square, c\cdot\square\bigr)$. This translates into finding rational numbers x, u, v and w with

$$\begin{aligned} x - e_1 &= a\,u^2\,;\\ x - e_2 &= b\,v^2\,;\\ x - e_3 &= c\,w^2\,. \end{aligned}$$

Eliminating x, this reduces to the two equations

$$a\,u^2 - b\,v^2 = e_2 - e_1 \quad \text{and} \quad a\,u^2 - c\,w^2 = e_3 - e_1. \tag{7.9}$$

Here the constants a, b, c, e_1, e_2, e_3 are given and we are looking for u, v, w in $\mathbb{Q}$. If we can prove that there cannot be a rational solution to these equation, then we can shrink the upper bound to r, otherwise if we can spot a solution we can find x and then P.

Example. Let us consider our example $y^2 = (x-1)(x-6)(x+7)$. We try to determine the image of $\kappa_1 \times \kappa_2$ in $H \times H$ where H is generated by -1, 2, 5 and 13. We already know that the image contains the subgroup generated by $(-1,-1)$, $(-10,10)$ and $(-2,13)$ which is a subspace of dimension 3 over $\mathbb{F}_2$ in a vector space $H \times H$ of dimension 8. Now we are using equation (7.9), which in our case is

$$a\,u^2 - b\,v^2 = 5 \quad \text{and} \quad a\,u^2 - ab\,w^2 = -8$$

where (a,b) runs through $H\times\ H$. Write the rational numbers with a common denominator $u = U/T$, $v = V/T$ and $w = W/T$. This turns it into equations to solve in integers U, V, W, T with no common divisor to all four integers:

$$a\,U^2 - b\,V^2 = 5\,T^2 \quad \text{and} \quad a\,U^2 - ab\,W^2 = -8\,T^2. \tag{7.10}$$

Consider the following grid representing the elements (a,b) in $H\times H$:

$a \setminus b$	1	−1	2	−2	5	−5	10	−10	13	−13	26	−26	65	−65	130	−130
1	O															
−1		P														
2									$P+T_3$							
−2										T_3						
5													T_2			
−5														$P+T_2$		
10					$P+T_1$											
−10						T_1										
13																
−13																
26																
−26																
65																
−65																
130																
−130																

Here are a few facts which are not too hard to prove: There are no solutions to (7.10) if

- $a < 0 < b$ or $b < 0 < a$;
- 13 divides a;
- $5 \mid a$ but $5 \nmid b$;
- $5 \mid b$ but $5 \nmid a$;
- b is even.

As example let us prove the middle statement: If $5 \mid a$ then $5 \mid aU^2 - 5T^2 = bV^2$. As $5 \nmid b$, we have $5 \mid V$. Also 5 divides $a(U^2 - bW^2) = 8T^2$ so $5 \mid T$. Hence 5^2 divides $5T^2 + bV^2 = aU^2$. As a is square-free it is not divisible by 25 which means that $5 \mid U$. Similarly $25 \mid abW^2$ implies that $5 \mid W$. We reached a contradiction as 5 divides all four variables.

Using the five conditions above, one can exclude many possibilities in the big grid. We are left with 16 elements in $H \times H$ that could belong to the image of κ. For 8 of them we know a point. But, for instance $(2, 1)$ is not yet excluded.

Let us prove that $(2,1)$ is not in the image of κ: The first equation $2U^2 - V^2 = 5T^2$ modulo 5 gives $2U^2 \equiv V^2 \pmod{5}$. As 2 is not a square modulo 5 this implies that $5 \mid U$ and $5 \mid V$. Now 5^2 divides $5T^2$ which shows that also T is divisible by 5. From the second equation we deduce the contradiction that 5 also divides W.

Therefore, we are left with 15 elements; however the number of elements in the image of κ is a power of 2. So it must be 8. We conclude that $E(\mathbb{Q})/2E(\mathbb{Q})$ is generated by T_1, T_2 and P and has 8 elements. This shows that $r = 1$. One could now use heights and prove that $E(\mathbb{Q})$ itself is generated by these three points. $\diamond$

In the example above, we were successful in finding the rank of $E(\mathbb{Q})$. In most examples with reasonably small coefficients this is the case. However there are situation when it is really difficult. One reason could be that the equation (7.9) are soluble, but all solutions have very large height. Another reason could be that the equations are a counter-example to the Hasse principle (as discussed in G12ALN=MATH2015); namely there are real solutions and there are solutions modulo m for all integers m, yet there are no integer solutions. In these cases it is very difficult to come up with a proof of the insolubility. The so-called Tate-Shafarevich group Ш$(E/\mathbb{Q})$ measures this.

Also, it turns out that it is more efficient to search for solutions for the equations (7.9) than for the original Weierstrass equation.

8 Additional topics

8.1 History

Disclaimer: I am not a historian, so this may not be as accurate as I wish. Mostly I learned about the history of elliptic curves from reading papers by Franz Lemmermayer, in particular from [5].

Elliptic curves occurred for the first time, if only implicitly, in the work of Diophantus of Alexandria, and the topic has remained close to the mathematical branch of diophantine geometry throughout the centuries. A typical example from Diophantus is the following: given a number, say 7, that can be written as a difference of two cubes ($7 = 8 - 1$), find two positive rational numbers a and b such that $7 = a^3 + b^3$. Diophantus succeeded with clever substitutions, and it was discovered much later by Isaac Newton (1643–1727), Édouard Lucas (1842–1891) and James Joseph Sylvester (1814–1897) that there is a geometric interpretation for these manipulations. In the example above, consider the curve $E : x^3 - y^3 = 7$ and the rational point $P = (2, 1)$. Then the chord and tangent construction will give a new rational points and some of them will have positive coordinates.

These kind of diophantine problems were a big thing for Pierre de Fermat (1607–1665) and Leonhard Euler (1707–1783), but then Carl Friedrich Gauss (1777–1855) started giving number theory a new direction by proving quadratic and biquadratic reciprocity laws. Generalising these to higher powers was a central occupation for a lot of number theorists.

In mathematical analysis certain inverse functions to elliptic integrals where studied by Niels Henrik Abel (1802–1829) in his proof that equations of degree larger than 4 cannot be solved by radicals. Carl Gustav Jacob Jacobi (1804–1851) and Karl Weierstrass (1815–1897) developed the theory of these elliptic functions further.

During these times, elliptic curves were studied mainly by less known number theorists like Augustin-Louis Cauchy (1789–1857), Lucas, Sylvester, Henri Poincaré (1854–1912) and Beppo Levi (1875–1961) (most of whom are known for their contributions to other areas of mathematics) as well as by what we would classify as complex algebraic geometers like Alfred Clebsch (1833–1872) or Christian Juel (1855–1935). Clebsch, in the 1860s, proved that curves of genus 0 are parametrized by rational functions, and that those of genus 1 are parametrized by elliptic functions. Juel was the first to point out the geometric interpretation of the group law in the 1890s.

Louis J. Mordell (1888–1972) proved a tacit assumption in one of Poincaré's articles, namely that the group of rational points on elliptic curves is generated by finitely many points. André Weil (1906–1998), one of the greatest mathematicians of the 20th century, gave a new and clearer proof of Mordell's theorem and generalised it to abelian varieties over number fields. He was also a pioneer in applying geometric methods to the study of algebraic equations over finite fields. He extended the work of Helmut Hasse (1898–1979) on the number of solutions of elliptic curves over finite fields and launched a far-reaching programme which was eventually completed by Pierre Deligne (1944–) in the 1970s.

It was, by the way, Weil's student Élisabeth Lutz (1914–2008) and, independently, Trygve Nagell (1895–1988) who first proved Theorem 6.7 on the torsion points. Another important result on the structure of torsion groups was obtained by Jean-Pierre Serre (1926–) in the early 1970s. This lead Barry Mazur (1937–) to prove Theorem 6.9 described earlier. There are still some open problems relating to torsion points on elliptic curves, like Serre's uniform boundedness conjecture.

The method of the proof of the Mordell-Weil Theorem 7.9 was refined by Ernst Sejersted Selmer (1920–2006) and Ian Cassels (1922–2015). John Tate (1925–) and Igor Shafarevich (1923–2017) used co-called Galois cohomology to describe the groups found by Selmer. In their honour, the group $Ш(E/\mathbb{Q})$ is called the Tate-Shafarevich group. It is a conjecture that this group is finite.

In the 1960s using one of the first computers in the world, Peter Swinnerton-Dyer (1927–) and Bryan Birch (1931–) discovered a conjecture which is still open. It was chosen among the seven millennium problems by the Clay Mathematics Institute. We will describe it in some details later.

There is a strong connection between elliptic curves over $\mathbb{Q}$ and modular forms. The work of Goro Shimura (1930–) allowed to obtain elliptic curves from certain modular forms. In the 1990s, it was the ground-breaking achievements of Andrew Wiles (1953–) and Richard Taylor (1962–) that showed that every elliptic curve can be obtained in this way. This was sufficient to conclude Fermat's Last Theorem.

Beyond the abstract interest in elliptic curve, there are now also important applications. Neil Koblitz (1948–) and Victor S. Miller (1947–) discovered in the 1980s that elliptic curves can be used in cryptography. In computational number theory, it was Hendrik Lenstra (1949–) who found a way to use elliptic curves for factorisation of large integers.

Research in elliptic curves is still a very active area. Major important results were recently found by Kazuya Kato, Richard Taylor, Christopher Skinner, Éric Urban, Manjul Bhargava, Wei Zhang, . . .

8.2 Cryptography

The basic idea behind modern asymmetric cryptosystems is the use of a trapdoor function: something that can be done easily but is very hard to undo. The examples discussed in G13CCR=MATH3011 are either built on the problem of factoring large integer or on the discrete logarithm problem. Elliptic curves over finite fields have their own discrete logarithm problem.

Let k be a finite field with a huge number of elements. You may think of $k = \mathbb{F}_p$ with p a prime number having hundreds or thousands of decimal digits. In practice one often uses certain fields $\mathbb{F}_{2^k}$ of characteristic 2 because they are more efficient to implement in a computer.

Let E be a fixed elliptic curve with coefficients in k and let $P \in E(k)$. We suppose that P has large order in $E(k)$. The data p, E, P can be made public and in fact they are often part of the standards implemented in a cryptographic library or a browser.

Note first that the multiplication of P by any large integer n can be done relatively fast using the following method. It is the elliptic curve analogue of fast modular exponentiation.

- Set $Q = O \in E(k)$ and $R = P$ and $k = n$.
- While $k > 0$ do the following:
  - ○ If k is odd replace Q by $Q + R$.
  - ○ Replace R by $2\,R$.
  - ○ Divide k without remainder by 2.
- Return Q

It takes at worst $\log_2(n)$ additions on the curve E and the same number of multiplications by 2. So even if n has hundreds of digits this can be done very quickly.

One central theorem on elliptic curves over finite fields was found by Hasse.

Theorem 8.1. *Let E/k be an elliptic curve over a finite field k. Then $|E(k)|$ is between $\#k - 2\sqrt{\#k} + 1$ and $\#k + 2\sqrt{\#k} + 1$.*

In average the number of elements in the group $E(k)$ is about $\#k = p$. For cryptography it is best to take $|E(k)|$ to be a prime number $N \neq p$ and P to be a generator of this cyclic group $E(k)$. We will from now on suppose that we are in this case, i.e., $E(k) \cong {}^{\mathbb{Z}}/_{N\mathbb{Z}}\, P$.

Discrete Logarithm Problem. Given a point $Q \in E(k)$, find $n \in {}^{\mathbb{Z}}/_{N\mathbb{Z}}$ such that $Q = n\,P$.

This is considered a computationally very hard problem. There are very few ideas how to compute the discrete logarithm problem. Baby-step-giant-step method for instance still require about $\sqrt{N}$ steps which is too much to do for a computer if N is large.

Now one can adapt the encryption, decryption, key exchange and signature methods to elliptic curves and get secure methods as long as nobody is able to break the discrete logarithm problem on that particular curve. As an example we will illustrate the Diffie-Hellman key exchange.

Suppose two people, usually called Alice and Bob, want to agree on a common secret key without wanting to meet physically. For instance such a key could then be used for a not so expensive symmetric encryption method like AES.

- They first agree on p, E and P.
- Alice chooses a random n and multiplies $Q = n\,P$ on the curve. Alice keeps n secret.
- Bob also choose a random m and multiplies $R = m\,P$ and keeps m secret.
- Now Alice send Q over to Bob and Bob send R over to Alice.
- Alice now multiplies R by n to get $n\,R = nmP$. Similar Bob computes $m\,Q = mn\,P = nm\,P$.

They both have the point $nm\,P$ whose x-coordinate they can use as a key for instance. An attacker listening to the conversation will know p, E, P, Q, R, but does not know n, m, or $nm\,P$. If they know how to break the discrete logarithm problem they can get the key.

Elliptic curve cryptography is now widely used. On the one hand the operations on the curve are more costly than the operations in $({}^{\mathbb{Z}}/_{p\mathbb{Z}})^{\times}$ used in classical cryptosystems. On the other hand the system are believed to be securer as the known methods for breaking the elliptic curve discrete logarithm are less efficient than the ones for the usual discrete logarithm. Therefore one can work with smaller p. It turns out in the end that elliptic curve cryptography is faster than the classical one. You probably use elliptic curve cryptography daily on your mobile phone or online on your computer.

8.3 Primality testing

Let n be a large integer. We already discussed in G12ALN=MATH2015 methods to check if n is composite without having to find factors. The idea is to use Fermat's Little Theorem. If $a^{n-1} \not\equiv 1 \pmod{n}$ then n is composite. We used fast modular exponentiation to evaluate a^{n-1} in about $\log_2(n)$ steps.

Suppose now n is a number that has passed a few of these tests and also we used trial division to be sure that no small prime divides it. The number n is either prime or one of the few numbers that are pseudo-prime to many bases. How could we now prove that n is prime?

Henry Cabourn Pocklington (a school teacher in Leeds, 1870–1952) found the following method.

Theorem 8.2. *Suppose $n-1 = \ell_1^{\alpha_1}\cdot\ell_2^{\alpha_2}\cdots\ell_s^{\alpha_s}\cdot r$ with primes ℓ_i and some $r<\sqrt{n}$ whose factorisation we do not know. If for each $1\leqslant i\leqslant s$ we find an integer a_i such that $a_i^{n-1}\equiv 1 \pmod{n}$ and n is coprime to $a_i^{(n-1)/\ell_i}-1$, then n is a prime number.*

This is very efficient in case we know how to find enough of the factorisation of $n-1$, but that may be very difficult.

Shafi Goldwasser (1959–) and Joe Kilian (1963–) proposed the following theorem that can be used very efficiently.

Theorem 8.3. *Let A and B be in $\mathbb{Z}/n\mathbb{Z}$ such that $E: y^2 = x^3 + A\,x + B$ has discriminant $\Delta = -16(4A^3 - 27B^2)$ coprime to n. Suppose there exist prime numbers $\ell_1, \ell_2, \dots, \ell_s$ and points $P_1, P_2, \dots, P_s$ in $E(\mathbb{Z}/n\mathbb{Z})$ of the form $P_i = (x_i : y_i : 1)$ such that*

- *$\ell_i\,P_i = O$ and*
- *$\prod_{i=1}^{s}\ell_i > \bigl(\sqrt[4]{n}+1\bigr)^2$.*

then n is a prime number.

Proof. Suppose p is a prime divisor of n. Consider the reduced points $\tilde{P}_i$ on the curve $\tilde{E}$ reduced modulo p. The assumptions guarantee that $\tilde{P}_i$ has order equal to ℓ_i in $\tilde{E}(\mathbb{F}_p)$. It follows that $\prod \ell_i$ divides the order of the group $\tilde{E}(\mathbb{F}_p)$. Therefore

$$\bigl(\sqrt[4]{n}+1\bigr)^2 < \prod_{i=1}^{s}\ell_i \leqslant \bigl|\tilde{E}(\mathbb{F}_p)\bigr|.$$

The theorem by Hasse 8.1 implies that $|\tilde{E}(\mathbb{F}_p)| < (\sqrt{p}+1)^2$. Therefore $\sqrt[4]{n} < \sqrt{p}$. But $\sqrt{n}<p$ is impossible unless $n=p$. Hence n is prime. □

Example. Suppose we want to test that $n = 907$ is prime. Pick $E : y^2 = x^3 + 10x - 2$ and $P = (819, 784)$. Then $71\,P = O$. Since 71 is prime and $71 > (1 + \sqrt[4]{907})^2 \approx 42.1$, we conclude that n is prime. $\diamond$

If we did not know in the above that 71 is a prime, we could use the method again with $n = 71$ and some other elliptic curve. In general, we may find a point P_i is of some order, which we suspect to be a prime ℓ_i. So we first have to use the algorithm to check that the much smaller prime ℓ_i is indeed a prime.

Example. Say we wish to certify that $n = 12345678901234567891$ is a prime number. We pick the curve $E : y^2 = x^3 - 1$ over ${}^{\mathbb{Z}}\!/_{n\mathbb{Z}}$ and find on it a point $P = (12080163983756654009, 5815984457902168698)$ whose order is $\ell = 253595994697$. This satisfies the theorem except that we are not certain that ℓ is prime.

Now we take the curve $E_2 : y^2 = x^2 + 2x + 1$ defined over ${}^{\mathbb{Z}}\!/_{\ell\mathbb{Z}}$. On it there is a point $(221240833535, 87782106347)$ having order $\ell_2 = 545387$. This again satisfies the conditions except that we have to certify that ℓ_2 is prime. At this stage this can be done quickly with trial division. This now shows that n is a prime number. $\diamond$

Of course, the efficiency of this method needs that one can multiply points on an elliptic curve even by huge integers in reasonable time and that we can find the order of a point quickly. The first is done with fast multiplication explained in the previous section. The latter can be computed fairly quickly with a method like baby-step-giant-step.

8.4 Factorisation

Given a large integer n of which we know that it is not a prime number, say because it did not pass a Fermat primality test. Our aim is to find a non-trivial divisor $1 < d < n$ of n.

The basic idea is the following. Just pretend that n is a prime. Pick an elliptic curve $E : y^2 = x^3 + A\,x + B$ with coefficients A and B in $\mathbb{Z}/n\mathbb{Z}$. Pick a point P with coordinates there. Now start multiplying the point and adding it to others. In doing so we have to compute expressions like

$$\frac{y_P - y_Q}{x_P - x_Q} \qquad \text{or} \qquad \frac{3x_P^2 + A}{2y_P}.$$

It can happen that a denominator D above is not invertible modulo n. Either it is a multiple of n, in which case it is of no use, or the greatest common divisor of n and D is a non-trivial divisor d.

The obvious first objection should be that it seems rather random and we might as well just pick random numbers and compute their greatest common divisor with n. That, however, would be very inefficient for large n. Instead the elliptic curve method is much more likely to find a divisor.

A prime number p is C-smooth if all prime factors ℓ of $p-1$ are smaller than C. Recall the classical Pollard's $p-1$ method for factorisation. Given a bound C set $K = C!$. Pick a random a and use fast modular exponentiation to evaluate a^K modulo n. If $a^K \not\equiv 1 \pmod{n}$ then it is likely that $\gcd(n, a^K - 1)$ is a non-trivial divisor of n.

The following is Lenstra's version using elliptic curves:

- Pick random A, x_0 and y_0 in $\mathbb{Z}/n\mathbb{Z}$.
- Set $B = y_0^2 - x_0^3 - A\,x_0$ and check that $\Delta = -16\,(4A^3 + 27B^2)$ is coprime to n.
- Let E be the elliptic curve $y^2 = x^3 + Ax + B$ and $P = (x_0, y_0) \in E(\mathbb{Z}/n\mathbb{Z})$.
- Choose a bound C, maybe a million or so.
- Compute $(C!) \cdot P$ by computing successively P, $2\,P$, $3 \cdot 2\,P$, $4 \cdot 3 \cdot 2\,P$, ...
- If at some stage this computation fails, it could be because a denominator has a non-trivial factor with n and we are done. Otherwise start again with a different choice of E and P.

Of course, we can do this in parallel on thousands of computers to increase our chances to find a divisor. In practise, this method is efficient for finding factors that have up to 30 decimal digits. For integers n with more than 60 digits other factorisation methods, like sieve methods, are more efficient.

We have yet to explain why the method works. Suppose p is a prime number dividing n. Let $\tilde{E}/\mathbb{F}_p$ be the equation $y^2 = x^3 + Ax + B$ with the coefficients reduced modulo p. Let $\tilde{P}$ be the reduced point in $\tilde{E}(\mathbb{F}_p)$. The order of the group $\tilde{E}(\mathbb{F}_p)$ is approximatively p by the theorem of Hasse 8.1. Most integers with 30 digits will be C-smooth numbers which implies that $(C!) \cdot \tilde{P} = O$. Hence at some stage $(m!) \cdot \tilde{P} = O$. This will mean that p will divide a denominator in the computation. It is unlikely that the $(m!) \cdot P$ reduced modulo a different prime factor q of n to O and so the appearing non-invertible denominator will not be divisible by q. So we are likely to find a non-trivial divisor of n.

A very much optimised implementation can be found at `http://ecm.gforge.inria.fr/`. Using this, my office computer can factor

$$602400691612422154516282778947806249229526581$$

in less than 0.2 seconds.

8.5 Elliptic curves over $\mathbb{C}$

We will start with something that does not seem to have to do anything with elliptic curves at all. But be patient, we will get there!

Let ω_1 and ω_2 be two complex numbers with different argument. Consider

$$\Lambda = \Bigl\{ a\,\omega_1 + b\,\omega_2 \Bigm| a, b \in \mathbb{Z} \Bigr\}$$

which is a subset of $\mathbb{C}$ called a lattice. It is naturally a subgroup of $\mathbb{C}$ with addition. The parallelogram

$$\mathcal{F} = \Bigl\{ x\,\omega_1 + y\,\omega_2 \Bigm| 0 \leqslant x < 1, 0 \leqslant y < 1 \Bigr\}$$

is called a fundamental parallelogram of Λ.

A meromorphic function $f \colon \mathbb{C} \to \mathbb{C} \cup \{\infty\}$ is called **doubly periodic** with period lattice Λ if $f(z+\omega) = f(z)$ for all $\omega \in \Lambda$. Meromorphic means roughly that, for all $a \in \mathbb{C}$, the function can be expanded as a Taylor series $a_r(z-a)^r + a_{r+1}(z-a)^{r+1} + \cdots$ for some $r \in \mathbb{Z}$ and $a_i \in \mathbb{C}$. Poles are the places where $r < 0$. Now complex function theory as in G12COF=MATH2007 shows that if f is doubly period but has no poles then it must be a constant function. Also the numbers of poles and zeros (counted with multiplicity) within $\mathcal{F}$ are equal. It is impossible to find such a function with a single simple pole in $\mathcal{F}$; the easiest doubly period function will have a pole of order two, meaning $r = -2$.

For a lattice Λ, define

$$\wp_\Lambda(z) = \frac{1}{z^2} + \sum_{\omega \in \Lambda \setminus \{0\}} \Bigl(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \Bigr).$$

It can be shown that this sum converges absolutely for all $z \in \mathbb{C}$. It gives a doubly periodic function with period lattice Λ having a single pole of order two at each point of the lattice. It is called the **Weierstrass $\wp$-function**. Automatically, its derivative

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

is also a doubly period function, but now with a single pole of order 3 in $\mathcal{F}$.

Theorem 8.4. *All doubly periodic functions for Λ can be obtained as quotients of polynomials in $\wp_\Lambda$ and $\wp'_\Lambda$.*

Other naturally defined doubly periodic functions are Jacobi's and Abel's elliptic functions. They appear when studying the generalisation of trigonometric functions from the circle to ellipses and when calculating arc lengths on the ellipse. Hence the name – which elliptic curves inherited.

The expansion of $\wp_\Lambda(z)$ at $z = 0$ looks like

$$\wp_\Lambda(z) = \frac{1}{z^2} + \frac{1}{20}\, g_4(\Lambda)\, z^2 + \frac{1}{28}\, g_6(\Lambda)\, z^4 + \cdots$$

where

$$g_4(\Lambda) = 60 \sum_{\omega\in\Lambda\setminus\{0\}} \frac{1}{\omega^4} \qquad \text{and} \qquad g_6(\Lambda) = 140 \sum_{\omega\in\Lambda\setminus\{0\}} \frac{1}{\omega^6}.$$

The functions $\wp_\Lambda$ and $\wp'_\Lambda$ satisfy the relation

$$(\wp'_\Lambda(z))^2 = 4\,\wp_\Lambda(z)^3 - g_4(\Lambda)\cdot\wp_\Lambda(z) - g_6(\Lambda).$$

This means that there is a map $z \mapsto (\wp_\Lambda(z), \wp'_\Lambda(z))$ whose image lands among the points of the curve

$$E_\Lambda : y^2 = 4\,x^3 - g_4(\Lambda)\,x - g_6(\Lambda).$$

This is an affine curve defined over $\mathbb{C}$ whose projective closure is a smooth projective curve with a natural point $(0 : 1 : 0)$ to make this into an elliptic curve. As the functions $\wp_\Lambda$ and $\wp'_\Lambda$ are doubly periodic, they induce a map

$$\varphi\colon {}^{\mathbb{C}}\!/_{\Lambda} \to E_\Lambda(\mathbb{C}).$$

Note the left hand side is an abelian group obtained as a quotient of $\langle\mathbb{C}, +\rangle$ by its subgroup Λ.

Theorem 8.5. *The map φ defines a group isomorphism from $\mathbb{C}/\Lambda$ to $E_\Lambda(\mathbb{C})$.*

How should we picture $\mathbb{C}/\Lambda$? Each coset of this quotient has exactly one element in the fundamental parallelogram F. So $\mathbb{C}/\Lambda$ looks topologically as $\mathcal{F}$ with the boundaries glues together in the natural way. Glueing two sides we get a cylinder and glueing the other two afterwards result in a torus, a bagel shaped surface.

One can show that every elliptic curve $E/\mathbb{C}$ is isomorphic to a E_Λ for some lattice Λ. We may change the basis elements ω_1 and ω_2 in Λ to $\omega'_1 = a\omega_1 + b\omega_2$ and $\omega'_2 = c\omega_1 + d\omega_2$ as long as $\gamma = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ belongs to $\mathrm{SL}_2(\mathbb{Z})$. Scaling and turning the lattice Λ will not change the elliptic curve so one can always change the lattice to the form $\Lambda_\tau = \mathbb{Z} + \mathbb{Z}\tau$ with $\omega_1 = 1$ and $\omega_2 = \tau$ for some $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$.

Hence every elliptic curve $E/\mathbb{C}$ is isomorphic to $E_\tau = E_{\mathbb{Z}+\mathbb{Z}\tau}$. Moreover two such E_τ and $E_{\tau'}$ are isomorphic if and only if $\tau' = (a\tau + b)/(c\tau + d)$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ as above. It is therefore interesting to study the upper half plane, that is all $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$, with its group action by $\mathrm{SL}_2(\mathbb{Z})$.

Functions g that associate (in an analytic way) to a lattice Λ a complex number $g(\Lambda)$ such that $g(\lambda\Lambda) = \lambda^{-k}\, g(\Lambda)$ for all $\lambda \in \mathbb{C}^\times$ are called modular forms of weight k and level 1. They are often written as a function in τ satisfying some functional equation with respect to all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$. The functions g_4 and g_6 and $\Delta = g_4^3 - 27 g_6^2$, the discriminant of E_τ, are modular forms of weight 4, 6 and 12 respectively.

It is very useful to write modular forms in terms of the new variable $q = \exp(2\pi i \tau)$. Then the discriminant has a very nice form

$$\Delta = q \cdot \prod_{n=1}^{\infty} (1 - q^n)^{24} = q - 24\, q^2 + 252\, q^3 - 1472\, q^4 + 4830\, q^5 + \cdots .$$

Now we have a nice justification why we multiplied our formula for Δ by -16.

8.6 Birch and Swinnerton-Dyer conjecture

How to win a million dollars with elliptic curves?

Let E be an elliptic curve defined over $\mathbb{Q}$. We saw in Section 7.3 that it can be very difficult to determine the rank of $E(\mathbb{Q})$. So even to know if an elliptic curve has infinitely many solutions or not can be very hard.

Birch and Swinnerton-Dyer experimented with specific curves on one of the world's first computers, the EDSAC2. They found an interesting way one might be able to the get rank r. The reasoning is the following: If there are plenty of points in $E(\mathbb{Q})$ then there should be a lot of points in $\tilde{E}(\mathbb{F}_p)$ for all primes p as we can reduce all of them modulo p. Let $N_p = |\tilde{E}(\mathbb{F}_p)|$. By Hasse's theorem 8.1 we know that N_p is about p. If $r > 0$ then we might expect that N_p is often larger than p.

They considered the function

$$f(X) = \prod_{\text{primes } p \leqslant X} \frac{N_p}{p}$$

as X is an increasing real number. Based on numerical experiments they conjectured the following.

Conjecture. $f(X)$ stays bounded if and only if there are only finitely many solutions in $\mathbb{Q}$.

Even further they conjectured that $f(X)$ grows like $\log(X)^r$, where r is the rank of $E(\mathbb{Q})$. This is still an open problem. But it is not in this form that it is useful as we have no means to actually determine the growth rate of $f(X)$ from a finite amount of data. Instead they reformulated it using the L-function of E, which we explain now.

Set $a_1 = 1$. If p is a prime of good reduction then set $a_p = p + 1 - |\tilde{E}(\mathbb{F}_p)|$. By Hasse's theorem 8.1, we have $|a_p| < 2\sqrt{p}$. For all $k > 1$, defined $a_{p^k} = p \cdot a_{p^{k-2}} - a_p \cdot a_{p^{k-1}}$. If p is a prime of bad reduction then set $a_{p^n} = \pm 1$ or 0 depending on a fixed and easy recipe. If n has prime factorisation $p_1^{k_1} \cdot p_2^{k_2} \cdots p_s^{k_s}$ then we define $a_n = a_{p_1^{k_1}} \cdots a_{p_s^{k_s}}$ so that $n \mapsto a_n$ is a multiplicative arithmetic function. Define the L-series attached to the elliptic curve E by

$$L(E, s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

It can also be written as an infinite product

$$L(E, s) = \prod_p \frac{1}{1 - a_p\, p^{-s} + \varepsilon(p)\, p^{1-2s}}$$

where $\varepsilon(p)$ is zero if the reduction is bad and $+1$ if the reduction is good. The two formulae should remind us of the Euler product formula for the Riemann zeta function. Hasse's theorem can be used to prove that $L(E, s)$ converges absolutely for $s > 3/2$.

If we plug $s = 1$ into the product (though we do not know that the result is meaningful as we have no way of showing that the product or the sum are convergent) we get

$$L(E, 1) \text{ “ = ” } \prod_p \frac{1}{1 - a_p/p + \varepsilon(p)/p} = \prod_p \frac{p}{N_p}$$

at least for good primes. This is a very informal link to the function $1/f(X)$ above.

Luckily since the work of Taylor and Wiles discussed in the next section there is a way to make sense of $L(E, s)$ for all $s \in \mathbb{R}$ and even $s \in \mathbb{C}$.

Conjecture. $L(E, 1) = 0$ if and only if $E(\mathbb{Q})$ is infinite. Furthermore, the order of vanishing of $L(E, s)$ at $s = 1$ is equal to the rank r of $E(\mathbb{Q})$

This means that the Taylor expansion of the function $L(E,s)$ at $s=1$ looks like $a_r(s-1)^r+\cdots$. There is even a precise conjectural formula what a_r should be. It is a bit complicated to explain, but it contains as one term the size of the mysterious group $Ш(E/\mathbb{Q})$.

This conjecture is an open problem and considered very difficult. The Clay Mathematics Institute has chosen it among the 7 millennium problems and promises a million dollars for the first person to prove (or disprove) this conjecture.

Some results are known about it. Here a result which was proven through work of Kolyvagin, Gross and Zagier and, in special cases, by Coates and Wiles.

Theorem 8.6. *If $L(E,1)\neq 0$ then $E(\mathbb{Q})$ is finite. If $L(E,1)=0$ and $L'(E,1)\neq 0$ then the rank r is 1.*

This confirms one implication for the cases of rank 0 and 1. The other implication is only known if we assume that $Ш(E/\mathbb{Q})$ is finite by the work of Zhang and Skinner. There are also a lot of partial results on the leading term formula. Yet, the conjecture is still open.

Note by the way, that the the initial reasoning is really too naive. Taylor showed that, for any elliptic curve, independent of the rank r, there are (in some precise sense) the same number of primes p such that N_p is larger than p as there are primes with N_p smaller than p. This is a consequence of the Sato-Tate distribution of the values a_p.

8.7 Fermat's Last Theorem

Theorem 8.7. *Let $n\geqslant 3$. Then the only points in $C(\mathbb{Q})$ for $C: X^n+Y^n=Z^n$ are those with $XYZ=0$.*

For $n=2$ the equation is a conic and $C(\mathbb{Q})$ is in bijection with $\mathbb{P}^1(\mathbb{Q})$. For $n=3$, the curve is an elliptic curve with precisely three points in $C(\mathbb{Q})$. The curve for $n=4$ is also linked to an elliptic curve.

So we may suppose that $n\geqslant 5$. Kummer already showed in the 19th century that the theorem is true for all $n\leqslant 100$ with the possible exception of 37, 59 and 67.

Here we illustrate vaguely the ideas that went into the proof finalised by Andrew Wiles in the 1990s.

Suppose $a^n+b^n=c^n$ is a solution to C with $abc\neq 0$. We may also suppose that they are pairwise coprime. Consider now the elliptic curve

$$E: y^2 = x(x-c^n)(x-b^n).$$

This is called a Frey curve; the aim is to show that it cannot exist. The discriminant of E is

$$\Delta = 16\bigl(c^n b^n (c^n - b^n)\bigr)^2 = 16\,(abc)^{2n}.$$

For all primes p not dividing $2abc$ the curve has good reduction. If p is odd and divides abc, then it divides exactly one of them. This shows that only one of $e_1 - e_2 = c^n$, $e_1 - e_3 = b^n$ and $e_2 - e_3 = a^n$ is divisible by p. It follows that the reduction of E at p is a nodal curve; so called multiplicative reduction.

Let ℓ be prime number not dividing any integer that appeared so far. Consider the points of order ℓ in $E(\mathbb{C})$. Their coordinates are in fact in the field of all algebraic numbers $\bar{\mathbb{Q}}$. For each prime $p \nmid abc\ell$ the map $E(\bar{\mathbb{Q}})[\ell] \to E(\bar{\mathbb{F}}_p)[\ell]$ is injective. The curve has such nice properties that (in some sense) even the map for $p \mid abc$ behaves well. In fact, it is the ℓ-adic Galois representation attached to E that has super good properties.

The main theorem of Taylor and Wiles shows that there exists a modular form f associated to E. We discussed modular forms quickly at the end of Section 8.5. What is this modular forms? Recall from the Section 8.6 that E has an L-series $L(E,s) = \sum_{n\geqslant 1} a_n n^{-s}$. The modular form $f(\tau)$ is simply $\sum_{n\geqslant 1} a_n q^n$ where $q = \exp(2\pi i\tau)$ is a function on $\tau \in \mathbb{C}$ with $\mathrm{Im}(\tau) > 0$. To say that this analytic function is a modular modular form forces it to satisfy certain precise functional equations involving a number called the level N. For our Frey curve the level is simply the product of all primes dividing abc. Ribet and Mazur proved earlier that if the Galois representation of f is super nice, then there is another modular form g with a lower level and the same Galois representation. The Frey curve would be so nice that this level would drop all the way down to $N = 2$. However it is not hard to show that there are no modular forms of that level. This yields the final contradiction on the assumption of the existence of the solution $(a : b : c)$.