Elliptic Curves

Summer semester 2023

Universität Bayreuth

Michael Stoll

Contents

Screen version of July 21, 2023, 14:00.

1. Introduction / Appetizer

In this introductory chapter, I would like to give a rough sketch of how one can use elliptic curves to factor large integers. Everything will be explained in detail later in the course.

For the purposes of this introduction, an *elliptic curve* E is simply an equation

$$E\colon y^2 = x^3 + a\,x + b \tag{1.1}$$

in the variables x and y with coefficients a and b from a field K (such that $\operatorname{char}(K) \neq 2$), with the additional condition that $4\,a^3 + 27\,b^2 \neq 0$; otherwise the curve is not "smooth". The we can define the *set of K-rational points on E*, written $E(K)$, as the set of solutions $(\xi, \eta) \in K \times K$ of Equation (1.1). There are good reasons (that will be explained soon) to add a further point O "at infinity" to this set. We therefore set

$$E(K) = \{(\xi, \eta) \in K \times K \mid \eta^2 = \xi^3 + a\,\xi + b\} \cup \{O\}\,.$$

What is the use of this? Apart from the fact that algebraic curves like E are an interesting object of study by themselves, the special property of elliptic curves is that their (rational) points form an *abelian group* in a natural way. This group structure can be defined very concisely in the following geometric way.

- O is the zero element, and
- the sum of three points that are on a line is zero.

One just has to be careful to count the intersection points of a line with E with the correct multiplicity (if the line is tangent to E at the point, the multiplicity is 2; if it is an inflectional tangent, it is 3) and that one has to consider O as the third point of intersection with any vertical line. This is very natural if we consider E as a *projective* curve. The geometric interpretation of the group law leads to the following formulas.

$$\begin{aligned}
-(\xi, \eta) &= (\xi, -\eta) \\
(\xi, \eta) + (\xi, -\eta) &= O \\
(\xi_1, \eta_1) + (\xi_2, \eta_2) &= (\xi_3, \eta_3) \qquad \text{with} \quad \xi_3 = \lambda^2 - \xi_1 - \xi_2, \quad \eta_3 = -\lambda\xi_3 - \mu,
\end{aligned}$$

where

$$\lambda = \begin{cases} \dfrac{\eta_2 - \eta_1}{\xi_2 - \xi_1}, & \text{falls } \xi_1 \neq \xi_2 \\[2ex] \dfrac{3\xi_1^2 + a}{2\eta_1}, & \text{falls } \xi_1 = \xi_2 \text{ and } \eta_1 \neq -\eta_2, \end{cases}$$

and $\mu = \eta_1 - \lambda\xi_1$; here $y = \lambda\,x + \mu$ is the equation of the line through the two points, or the tangent to E in the point when the two points coincide.

These formulas may look a bit complicated at first sight, but they show very clearly that one can easily do computations in this group. (As an aside, it is a fairly thankless task to check that this addition is associative just using these formulas. There are better ways of doing this.)

1.1. Example. We consider the curve

EX
Addition of points

$$E\colon y^2 = x^3 - 43\,x + 166\,.$$

It has the rational point $P = (3, 8) \in E(\mathbb{Q})$. We compute

$$2 \cdot P = (-5, -16), \quad 3 \cdot P = P + 2 \cdot P = (11, -32), \quad 4 \cdot P = (11, 32) = -3 \cdot P\,.$$

This implies that $7 \cdot P = O$. (It is a fact that $E(\mathbb{Q}) \cong \mathbb{Z}/7\mathbb{Z}$, generated by P. In general, $E(\mathbb{Q})$ does not have to be finite, but it is always finitely generated (Mordell's Theorem). We will discuss elliptic curves over $\mathbb{Q}$ in more detail later in this course.) ♣

How can we use this property to factor numbers? For this we need to consider the case that the base field K is a finite field $\mathbb{F}_p$. In this case the group $E(\mathbb{F}_p)$ is clearly also finite. One even knows fairly precisely how large it is: we have $\#E(\mathbb{F}_p) = p + 1 - t$ with $|t| \le 2\sqrt{p}$. (The heuristics behind this is that for a random $\alpha \in \mathbb{F}_p$, the equation $y^2 = \alpha$ as on average one solution, so thinking of the right hand side of the curve equation as something random, we expect $p + 1$ points plus some smallish deviation (the "$+1$" is for the point O).)

1.2. **Example.** We have the following table for the sizes $\#E_a^{\pm}(\mathbb{F}_{23})$, where for $a \in \mathbb{F}_{23}$ we consider $E_a^{\pm} \colon y^2 = x^3 \pm x + a$. (A dash stands for a singular curve.) **EX** $\#E(\mathbb{F}_p)$

a	0	1	2	3	4	5	6	7	8	9	10	11
$\#E_a^+$	24	28	24	27	29	22	21	18	28	20	32	33
$\#E_a^-$	24	–	30	30	31	18	22	28	21	32	23	25

a	12	13	14	15	16	17	18	19	20	21	22	
$\#E_a^+$	15	16	28	20	30	27	26	19	21	24	20	
$\#E_a^-$	23	25	16	27	20	26	30	17	18	18	–	

In addition, there are the two curves $y^2 = x^3 \pm 1$ with 24 points each. One can show that every elliptic curve over $\mathbb{F}_{23}$ occurs exactly once "up to isomorphism" in this list.

We have $|t| \le \lfloor 2\sqrt{23} \rfloor = 9$, and we obtain the following frequency distribution.

t	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9
	1	2	1	4	1	4	3	2	2	6	2	2	3	4	1	4	1	2	1

We see that all possibilities do occur, and that the distribution is roughly uniform. ♣

It is no coincidence that the frequency table for t above is symmetric about zero. We assume that p is odd. Fix a non-square $d \in \mathbb{F}_p^\times$. Then the map

$$E \colon y^2 = x^3 + ax + b \quad \longmapsto \quad E' \colon y^2 = x^3 + d^2ax + d^3b$$

induces an involution on the set of isomorphism classes of elliptic curves over $\mathbb{F}_p$, and one can show fairly easily that

$$\#E(\mathbb{F}_p) = p + 1 - t \quad \Longrightarrow \quad \#E'(\mathbb{F}_p) = p + 1 + t\,.$$

This implies that the number of isomorphism classes with $t = c$ is the same as the number of isomorphism classes with $t = -c$.

How can we now use elliptic curves for factoring? We first consider another method, which has inspired the approach using elliptic curves. This is "Pollard's $(p-1)$-method".

Let N be a (large) composite number that is not a prime power (and so N has at least two distinct prime factors). Both properties can be verified easily without having to factor N first. We want to find a nontrivial divisor d of N. To do this, we choose a random number $a \in \{2, \ldots, N-1\}$. If $d = \gcd(a, N) > 1$, then d is a nontrivial divisor of N, and we are done. Otherwise a is invertible mod N. We choose a number $L \ge 1$ and set $B = \operatorname{lcm}(1, 2, \ldots, L)$. Then we compute $d =$

$\gcd(a^B - 1, N)$. To do so, we first compute $b = a^B \bmod N$ by successive squaring and then $d = \gcd(b - 1, N)$. This requires $O((\log B)M(N))$ bit operations, where $M(N)$ is the complexity of a multiplication in $\mathbb{Z}/N\mathbb{Z}$. (The "naive" method gives $M(N) \in O((\log N)^2)$, asymptotically more efficient methods lead to $M(N) \in O((\log N)(\log\log N)(\log\log\log N))$.) The Prime Number Theorem implies that $\log B \approx L$. If $1 < d < N$, then we have found the desired factor of N.

Under which conditions can we hope to find a factor? This is likely to be the case when there are prime divisors p and q of N such that $p - 1 \mid B$ (this means that every prime power dividing $p - 1$ is $\le L$), but $q - 1 \nmid B$. Then $a^B - 1$ is divisible by p (since $a^{p-1} \equiv 1 \bmod p$ by Fermat's little theorem). On the other hand, $a^B - 1$ is quite unlikely to be divisible by q (otherwise, a has to be a kth power mod q, where $k = (q - 1)/\gcd(B, q - 1) \ge 2$; the probability for this is about $1/k$). So when we try several choices for a, then we will very quickly hit one for which the method works: we can expect that $d = \gcd(a^B - 1, N)$ is divisible by p, but not by q, and so d is a nontrivial divisor of N.

In practice, it is a good idea to use sequence of values for B that are obtained by successively multiplying

$$2 \cdot 3 \cdot 2 \cdot 5 \cdot 7 \cdot 2 \cdot 3 \cdot 11 \cdot 13 \cdot 2 \cdot 17 \cdot 19 \cdot 5 \cdot 3 \cdot 29 \cdot 31 \cdot \cdots ;$$

the sequence of factors comes from the sequence of prime powers

$$2,\ 3,\ 2^2,\ 5,\ 7,\ 2^3,\ 3^2,\ 11,\ 13,\ 2^4,\ 17,\ 19,\ 5^2,\ 3^3,\ 29,\ 31, \ldots .$$

The main problem with this method is that it can only work when N has prime divisors with the relevant properties.

H.W. Lenstra (*1949) Foto © MFO

At this point, elliptic curves come to the rescue. Hendrik Lenstra had the idea to replace the multiplicative group that we have used above by the group of points on an elliptic curve over $\mathbb{Z}/N\mathbb{Z}$. This provides us with a rich choice of different groups, and so we can hope to find one fairly quickly for which the order over $\mathbb{F}_p$ is "L-smooth" in that all prime power divisors of it are $\le L$, but the order over $\mathbb{F}_q$ is not. We therefore choose a random elliptic curve E with coefficients $a, b \in \mathbb{Z}/N\mathbb{Z}$, together with a point $P = (\xi, \eta) \in E(\mathbb{Z}/N\mathbb{Z})$. For example, we can pick a randomly and take

$$E\colon y^2 = x^3 + ax - a\,, \qquad P = (1, 1)\,.$$

We can consider E and P also with coefficients/coordinates in $\mathbb{F}_p$; we write $\tilde{E}$ and $\tilde{P}$ in this case. Assuming that $\#\tilde{E}(\mathbb{F}_p) = p + 1 - t$, we then have $(p + 1 - t) \cdot \tilde{P} = \tilde{O}$. Similarly to what we did in Pollard's method, we multiply P by $B = \operatorname{lcm}(1, 2, \ldots, L)$. If $p + 1 - t$ divides B, then $B \cdot \tilde{P} = \tilde{O}$. But usually, it will not be the case that $B \cdot P = O$. This will lead to a step in the algorithm, where we would have to perform a division in $\mathbb{Z}/N\mathbb{Z}$ be a nonzero element that is however not invertible. The gcd computation involved in this division will then provide us with a nontrivial divisor of N (in most cases, this will be p).

For this method to work well in practice, we need to have a good chance of choosing B not too large so that $m = \#\tilde{E}(\mathbb{F}_p)$ is a divisor of B (for the smallest prime factor p of N and a reasonably large fraction of elliptic curves over $\mathbb{F}_p$). One can in fact show that if one chooses L (and therefore B) in an optimal way, one obtains an algorithm whose (expected) running time is bounded by

$$Ce^{(\sqrt{2}+o(1))\sqrt{(\log p)(\log\log p)}}\,;$$

this means that the algorithm has *subexponential* complexity. Here p is the smallest prime divisor of N, and $o(1)$ represents a function that tends to zero as $p \to \infty$.

EX Factorization

1.3. **Example.** As a baby example we factor the number $N = 851$. We take the elliptic curve $E\colon y^2 = x^3 + 9\,x - 9$ over $\mathbb{Z}/851\mathbb{Z}$ with the point $P = (1, 1)$. To compute $B \cdot P$, we successively determine $P_0 = P$, $P_1 = 2 \cdot P_0$, $P_2 = 3 \cdot P_1$, $P_3 = 2 \cdot P_2$, $P_4 = 5 \cdot P_3$ and so on. In this way, we gather the least common multiples of the first natural numbers. Now for the computation proper:

(1) $P_1 = 2 \cdot P_0$:
We find $\lambda = 6, \mu = 846$, so $P_1 = (34, 652)$.

(2) $P_2 = 3 \cdot P_1$:
First $Q = 2 \cdot P_1$. We find $\lambda = 374, \mu = 701$, so $Q = (244, 802)$. Now $P_2 = P_1 + Q$. We find $\lambda = 487, \mu = 263$ and so $P_2 = (313, 486)$.

(3) $P_3 = 2 \cdot P_2$:
$\lambda = 502, \mu = 795$, so $P_3 = (333, 537)$.

(4) $P_4 = 5 \cdot P_3$:
First $Q_1 = 2 \cdot P_3$: $\lambda = 305, \mu = 241$ and $Q_1 = (451, 66)$.
Then $Q_2 = 2 \cdot Q_1$: $\lambda = 832, \mu = 125$ and $Q_2 = (310, 659)$.
Finally $P_4 = P_3 + Q_2$. The denominator of the expression for λ comes out as 23, which is not invertible. Therefore $23 = \gcd(851, 23)$ is a nontrivial divisor, and we have found the factorization $851 = 23 \cdot 37$.

What happens in the background is that the point P has order 10 in $E(\mathbb{F}_{23})$, so $P_4 = O$ in this group. On the other hand, P has order 29 in $E(\mathbb{F}_{37})$, so $P_4 \neq O$ in that group. ♣

2. Affine plane curves

Elliptic curves are special plane algebraic curves. We therefore need to get acquainted with these first, even though this necessitates the introduction of many new notions. For reasons of time, we cannot really go deeply into *Algebraic Geometry*, which is the part of mathematics that studies these and more general objects.

Naively, an *affine plane curve* describes the set of points of the plane whose coordinates satisfy a polynomial equation in two variable. To formalize this idea, we first need to describe the plane which our curves live in.

Here and in the following, K is an (arbitrary) field. We also fix an algebraic closure $\bar{K}$ of K. This field K is our *base field*; it is the home of the coefficients of our equations and (usually) of the coordinates of the points that we consider.

2.1. **Definition.** The *affine plane* $\mathbb{A}^2_K$ over K has the following properties.

DEF
affine plane

(1) For every extension field $L \supset K$ we have the *set of L-rational points* of $\mathbb{A}^2_K$, which is the set

$$\mathbb{A}^2_K(L) = \{(\xi, \eta) \mid \xi, \eta \in L\} = L \times L\,.$$

(2) A *regular function* on $\mathbb{A}^2_K$ is given by a polynomial $f \in K[x, y]$. For every extension field $L \supset K$, f defines (via substitution of the coordinates) a function

$$f_L \colon \mathbb{A}^2_K(L) \longrightarrow L\,, \qquad (\xi, \eta) \longmapsto f(\xi, \eta)\,.$$

The function $f_{\bar{K}}$ determines f uniquely.

The ring of regular functions $K[x, y]$ on $\mathbb{A}^2_K$ is also called the *affine coordinate ring* of $\mathbb{A}^2_K$ and is written $K[\mathbb{A}^2_K]$.

(3) A *rational function* on $\mathbb{A}^2_K$ is given by an element $f = g/h \in K(x, y)$ of the field of fractions $K(x, y)$ of $K[x, y]$.

f is *regular* at the point $P = (\xi, \eta) \in \mathbb{A}^2_K(L)$ if $h(\xi, \eta) \neq 0$. For each $L \supset K$, f defines a function

$$f_L \colon \{P \in \mathbb{A}^2_K(L) \mid f \text{ regular at } P\} \longrightarrow L\,.$$

f is again uniquely determined by $f_{\bar{K}}$.

The regular functions are then exactly the rational functions that are everywhere (i.e., on $\mathbb{A}^2_K(L)$ for all L) regular.

The field $K(x, y)$ of rational functions on $\mathbb{A}^2_K$ is also called the *function field* of $\mathbb{A}^2_K$ and is written $K(\mathbb{A}^2_K)$. ♢

This is an *operational* definition, i.e., it does not so much say what $\mathbb{A}^2_K$ "is", but what one can do with it. If you do not like this point of view, you can as a first approximation think of $\mathbb{A}^2_K$ as the association $L \mapsto L \times L$ that sends an extension field L to the set of L-rational points. The regular and rational functions are an important part of the picture, however (like the differentiable, holomorphic or meromorphic functions in analysis). If one wants to do it really "right" (like in modern algebraic geometry), one defines objects like $\mathbb{A}^2_K$ as "ringed spaces" that carry both structures. (In classical algebraic geometry the base field K is algebraically closed (or even fixed to be $\mathbb{C}$); then one can get by with identifying an object like the affine plane with the set of its (K-rational) points. This does not make sense anymore when working over an arbitrary field K.)

REM n-dim. affine space

2.2. **Remark.** In a completely analogous way, one defines $\mathbb{A}^n_K$, the *n-dimensional affine space over K.* ♠

DEF affine plane curve

2.3. **Definition.** An *affine plane curve* C over K is given by a non-constant polynomial $f \in K[x, y]$. We write $C\colon f(x, y) = 0$.

(1) For every extension field $L \supset K$, we have the *set of L-rational points* on C,

$$C(L) = \{P \in \mathbb{A}^2_K(L) \mid f_L(P) = 0\} = \{(\xi, \eta) \in L \times L \mid f(\xi, \eta) = 0\}\,.$$

(2) A *regular function* on C is an equivalence class of polynomials in $K[x, y]$, where two polynomials are equivalent if their difference is divisible by f. If g is a representative of such an equivalence class, then we obtain functions

$$g_L\colon C(L) \ni (\xi, \eta) \mapsto g(\xi, \eta) \in L\,,$$

that depend only on the equivalence class (since $f_L = 0$ on C).

The regular functions on C form a ring, the *affine coordinate ring* $K[C]$ of C. It is isomorphic to $K[x, y]/K[x, y] \cdot f$.

(3) A *rational function* on C is an equivalence class of rational functions $g/h \in K(x, y)$ such that f and h have no non-constant common divisor. Here g_1/h_1 and g_2/h_2 are equivalent if f divides $g_1h_2 - g_2h_1$.

A rational function ϕ is *regular* at a point $P \in C(L)$, if there is a representative g/h of ϕ with $h_L(P) \neq 0$. We then obtain for every $L \supset K$ a function

$$\phi_L\colon \{P \in C(L) \mid g/h \text{ regular at } P\} \longrightarrow L\,, \qquad P \longmapsto \frac{g_L(P)}{h_L(P)}\,.$$

(4) C is *irreducible*, if f is irreducible. C is *geometrically irreducible*, if f is absolutely irreducible (i.e. irreducible in $\bar{K}[x, y]$).

When C is irreducible, then $K[x, y] \cdot f$ is a prime ideal, and so the coordinate ring $K[C]$ is an integral domain. The rational function on C then form the field of fractions of $K[C]$, the *function field* $K(C)$ of C. ♢

The condition on common divisors in the definition of rational functions on C ensures that such a function is regular on C at all but finitely many points.

EX affine plane curves

2.4. **Examples.**

(1) A trivial example is the "x-axis" $C\colon y = 0$. Here $f = y$, and the rational points are $C(L) = L \times \{0\}$. The coordinate ring is $K[C] = K[x, y]/K[x, y]\cdot y \cong K[x]$, and the function field is $K(C) \cong K(x)$.

(2) A more interesting example is the "unit circle" $C\colon x^2 + y^2 = 1$ (with $f = x^2 + y^2 - 1$). We assume that $\operatorname{char}(K) \neq 2$. For every L we have at least the four rational points $(0, \pm 1)$ and $(\pm 1, 0)$, but there are usually more. One can show that

$$C(L) = \Bigl\{\Bigl(\frac{2t}{1+t^2}, \frac{1-t^2}{1+t^2}\Bigr) \Bigm| t \in L, t^2 \neq -1\Bigr\} \cup \{(0, -1)\}\,.$$

As an example of a rational function on C, we consider $g = \frac{y-1}{x}$. At which points is g regular? Clearly, g is regular at all points whose x-coordinate does not vanish, so in all points except possibly $(0, \pm 1)$. What about these two points? At $(0, -1)$ the denominator of g vanishes, but the numerator has the nonzero value -2, which implies that g cannot be regular there (otherwise $y - 1 = x\frac{y-1}{x}$ would have to have the value 0 at this point). At $(0, 1)$, on the other hand, both numerator and denominator of g vanish. We can move to another representative

$$\frac{y-1}{x} = \frac{(y-1)(y+1)}{x(y+1)} = \frac{y^2-1}{x(y+1)} \sim \frac{-x^2}{x(y+1)} = -\frac{x}{y+1}\,,$$

that is defined at $(0,1)$ (and gives the value 0 there). So $(0,-1)$ is the only point at which g is not regular.

(3) Every curve $C\colon y^2 = x^3 + ax + b$ is geometrically irreducible. This is because every nontrivial factorization of $f = y^2 - x^3 - ax - b$ would have to be of the form $(y - h_1(x))(y - h_2(x))$, which implies $h_2 = -h_1$ and then $x^3 + ax + b = h_1(x)^2$. The latter is impossible, since the left hand side has degree 3, whereas the right hand side has even degree. ♣

3. Projective plane curves

The affine plane and affine plane curves are fairly concrete objects that one can easily visualize (at least when $K = \mathbb{R}$ or K is contained in $\mathbb{R}$), but they have some disadvantages. If we take $K = \mathbb{C}$ (then what we do is close to complex analysis), then we see in examples that the point sets $\mathbb{C}^2$ or $C(\mathbb{C})$ are not compact. This means that they are, in a certain sense, "open", that something is "missing". In many cases one can already see this in the real picture, e.g., for a line, a parabola or a hyperbola (for an ellipse, one has to consider it over $\mathbb{C}$).

One consequence of this imperfection is that there are exceptions and special cases that one has to consider. The prototypical example is that two distinct lines always meet in exactly one point—unless they are parallel. To get rid of this annoying exception, one adds points to the affine plane. More precisely, for every family of parallel lines (corresponding to a "direction") there is a new point that is contained in exactly these lines. All these new points together form a new line, the so-called "line at infinity". Then it is true without exception that any two distinct lines meet in exactly one point and that there is exactly one line that passes through any two distinct given points.

We will now define this projective plane formally as an object of algebraic geometry. Note that this definition is more symmetric than what is hinted at in the discussion above. In fact, it is completely arbitrary which line is designated to be "the" line at infinity.

3.1. **Definition.** The *projective plane* $\mathbb{P}^2_K$ over K has the following properties.

DEF
projective plane

(1) For every extension field $L \supset K$, we have the *set of L-rational points of* $\mathbb{P}^1K$,

$$\mathbb{P}^2_K(L) = \{(\xi, \eta, \zeta) \in L^3 \mid (\xi, \eta, \zeta) \neq (0, 0, 0)\} / \sim_L ,$$

where the equivalence relation $\sim_L$ is given by

$$(\xi, \eta, \zeta) \sim_L (\xi', \eta', \zeta') \iff \exists \lambda \in L^\times : \xi' = \lambda\xi, \eta' = \lambda\eta, \zeta' = \lambda\zeta .$$

So the coordinates are defined only up to scaling.

The point that is represented by (ξ, η, ζ) is written $(\xi : \eta : \zeta)$.

According to this definition, one can also think of the points of the projective plane the the lines through the origin in three-dimensional affine space (the non-zero points on such a line correspond to all representative coordinate triples of the projective point). One recovers the affine plane by identifying it with the plane $z = 1$ in this affine space: the lines through the origin that are not contained in the xy plane, meet the plane $z = 1$ in a unique point; this gives an embedding of $\mathbb{A}^2_K$ into $\mathbb{P}^2_K$. The remaining lines correspond to the points at infinity given by their direction. In formulas, the embedding reads

$$\mathbb{A}^2_K(L) \ni (\xi, \eta) \mapsto (\xi : \eta : 1) \in \mathbb{P}^2_K(L) ;$$

its inverse is defined for the points whose Z-coordinate does not vanish (this is independent of scaling) and is given by $(\xi : \eta : \zeta) \mapsto (\xi/\zeta, \eta/\zeta)$. The remaining points (with $Z = 0$) are exactly the L-rational points of the "line at infinity" $Z = 0$ (see below).

(2) Recall that a polynomial $f \in K[X, Y, Z]$ is *homogeneous* of degree d, if it has the form

$$f = \sum_{r+s+t=d} a_{rst} X^r Y^s Z^t .$$

A *rational function* on $\mathbb{P}^2_K$ is given by an element $f/g \in K(X,Y,Z)$, where f and g area homogeneous polynomials of the same degree.

f/g is *regular* at $P = (\xi : \eta : \zeta) \in \mathbb{P}^2_K(L)$, if $g(\xi,\eta,\zeta) \neq 0$ (since g is homogeneous, this condition does not depend on the scaling!). We obtain functions

$$(f/g)_L \colon \{P \in \mathbb{P}^2_K(L) \mid f/g \text{ regular at } P\} \ni (\xi : \eta : \zeta) \mapsto \frac{f(\xi,\eta,\zeta)}{g(\xi,\eta,\zeta)} \in L\,.$$

Note that this is well-defined, since f and g are both homogeneous and have the same degree. ♢

In contrast to the affine plane, there are no non-constant regular functions on the whole projective plane: a non-constant polynomial does not define a well-defined function, and a quotient f/g always has points in $\mathbb{P}^2_K(\bar{K})$ where g vanishes.

3.2. **Remark.** One can again define in an analogous way the *n-dimensional projective space $\mathbb{P}^n_K$ over K*. $\mathbb{P}^1_K$ is also called the *projective line* over K. ♠

REM
n-dim. proj. space

Projective plane curves are defined essentially analogously to affine plane curves. We just have to take care that the polynomial equation defining the curve gives a well-defined condition. This is achieved by using homogeneous polynomials.

3.3. **Definition.** A *projective plane curve C of degree $d \geq 1$* over K is given by a homogeneous polynomial $0 \neq f \in K[X,Y,Z]$ of degree d. (We write $C \colon f(X,Y,Z) = 0$.)

DEF
projective plane curve

(1) For any extension field $L \supset K$, we have the *set of L-rational points on C*,

$$C(L) = \{(\xi : \eta : \zeta) \in \mathbb{P}^2_K(L) \mid f(\xi,\eta,\zeta) = 0\}\,.$$

(2) A *rational function* on C is an equivalence class of rational functions on $\mathbb{P}^2_K$, whose denominator has no non-constant common divisor with f. Here g_1/h_1 and g_2/h_2 are equivalent if and only if f divides $g_1h_2 - g_2h_1$.

A rational function ϕ is *regular* at $P \in C(L)$, if it has a representative g/h such that h does not vanish in P. We then again obtain functions

$$\phi_L \colon \{P \in C(L) \mid g/h \text{ regular at } P\} \longrightarrow L\,.$$

(3) C is *irreducible*, if f is irreducible (in $K[X,Y,Z]$). C is *geometrically irreducible*, if f is absolutely irreducible.

If C is irreducible, then the rational functions on C form again a field, the *function field* $K(C)$ of C. ♢

It is now easy to go back and forth between "affine" and "projective".

First consider an affine curve $C \colon f(x,y) = 0$, and let d be the (total) degree of the defining polynomial f. Then $F(X,Y,Z) = Z^d f(X/Z, Y/Z)$ is a homogeneous polynomial of degree d (it is obtained from f by replacing x by X and y by Y and then by multiplying each monomial by a power of Z so that the total degree becomes d). The projective curve $\bar{C} \colon F(X,Y,Z) = 0$ is called the *projective closure* of C; the "extra points" in $\bar{C}(L) \setminus C(L)$ (these are the points with zero Z-coordinate) are the *points at infinity* of C or on $\bar{C}$.

Conversely, if $C \colon F(X,Y,Z) = 0$ is a projective curve of degree d, then $f(x,y) = F(x,y,1)$ is a polynomial of degree at most d, and the affine curve $C' \colon f(x,y) = 0$ is an *affine patch* of C (we obtain other affine patches by setting X or Y equal to 1). If $F = aZ^d$, we get the constant polynomial $f = a$, which does not define an affine curve, however. In this case, C has points only on the line at infinity.

These operations are essentially inverses of each other. The (standard) affine patch of the projective closure of an affine curve C is again C. Conversely, the projective closure of the standard affine patch of a projective curve C is again C, if the defining polynomial F is not divisible by Z (i.e., the line at infinity is not contained in C).

3.4. **Examples.**

EX
projective
closure

(1) The projective closure of the affine line $ax + by = c$ is the projective line $aX + bY - cZ = 0$. It has precisely one point $(-b : a : 0)$ at infinity. All projective lines (in $\mathbb{P}^2_K$) are obtained in this way, except the line $Z = 0$ "at infinity" (which consists of the points at infinity).

(2) The projective closure of the unit circle $x^2 + y^2 = 1$ is $X^2 + Y^2 - Z^2 = 0$. It has the two L-rational points at infinity $(1 : \pm i : 0)$, if $-1 = i^2$ is a square in L (and $\operatorname{char}(L) \neq 2$, otherwise one has the single point $(1 : 1 : 0)$). More generally, all circles $(x-a)^2+(y-b)^2 = r^2$ have the same two points at infinity.

(3) The projective closure of the affine curve $y^2 = x^3 + ax + b$ is

$$Y^2Z - X^3 - aXZ^2 - bZ^3 = 0\,.$$

It has exactly one (always rational) point $(0 : 1 : 0)$ at infinity. ♣

4. Intersections of curves and lines

In this section, we want to show that a projective line and a projective curve of degree d always intersect in exactly d points. We will need this result for the definition of the group structure on an elliptic curve. For the statement to be correct, we have to count the intersection points with the correct multiplicity. Therefore we first need to define this multiplicity.

In the following K is always our base field and L is a field extension of K.

4.1. **Definition.** Let $P = (\xi : \eta : \zeta) \in \mathbb{P}^2_K(L)$ be a point, $G\colon aX + bY + cZ = 0$ a projective line over K, and $C\colon F(X,Y,Z) = 0$ a projective curve over K. We assume that $aX + bY + cZ$ does not divide F (otherwise G would be contained in C). We define $i(G,C;P)$, the *multiplicity of the intersection point P of G and C*, as follows.

DEF
Intersection multiplicity

In the case that $P \notin C(L) \cap G(L)$ (so P is not a point of intersection of C and G), we set $i(G,C;P) = 0$. In the other case we solve the equation of G for one of the variables, e.g., $Z = -\frac{a}{c}X - \frac{b}{c}Y$ (if $c \neq 0$), and plug this expression into F. We obtain a homogeneous polynomial H in two variables that is divisible by $(\xi Y - \eta X)$ (if we eliminated Z, else by $(\xi Z - \zeta X)$ or $(\eta Z - \zeta Y)$). The multiplicity of this factor in H is then $i(G,C;P)$. ♢

The definition is of course independent of the choice of variable that is eliminated (exercise).

4.2. **Example.** We consider the curve $C\colon Y^2Z - X^3 + XZ^2 = 0$. For the line $Y = 0$ we obtain $H = -X^3 + XZ^2 = X(X+Z)(-X+Z)$; we therfore have multiplicity 1 in each of the three intersection points $(0:0:1)$, $(-1:0:1)$, and $(1:0:1)$.

EX
multiplicity of intersection points

For the line $X - Z = 0$ we obtain the following. We eliminate Z to get $H = XY^2$, so the intersection point $(1:0:1)$ has multiplicity 2. (Indeed, the line is the tangent line to the curve in this point.) The remaining intersection point $(0:1:0)$, on the other hand, has multiplicity 1.

We finally consider the line $Z = 0$. In this case we find $H = -X^3$, and so we have an intersection point of multiplicity 3 at $(0:1:0)$. (Here the line is the inflectional tangent.) ♣

This example already indicates that and why the following theorem is correct.

4.3. **Theorem.** *Let $C\colon F(X,Y,Z) = 0$ be a projective curve of degree d over K, and let $G\colon aX + bY + cZ = 0$ be a projective line over K that is not contained in C. Then*

$$\sum_{P \in C(\bar{K}) \cap G(\bar{K})} i(G,C;P) = d\,.$$

THM
Bézout's Theorem (special case)

If $L \supset K$ is an extension field such that

$$\sum_{P \in C(L) \cap G(L)} i(G,C;P) \geq d - 1\,,$$

then we have in fact that

$$\sum_{P \in C(L) \cap G(L)} i(G,C;P) = d\,.$$

The last statement says that the last intersection point is also L-rational, if this is true for all the others.

Proof. W.l.o.g., $c \neq 0$. We set $a' = -a/c$, $b' = -b/c$; then the equation of the line is $Z = a'X + b'Y$. We plug this into F and obtain $H(X, Y) = F(X, Y, a'X + b'Y)$; this is a homogeneous polynomial of degree d in $K[X, Y]$. As such, it splits in $\bar{K}[X, Y]$ into linear factors,

$$H(X, Y) = \alpha(\eta_1 X - \xi_1 Y)^{d_1} \dots (\eta_k X - \xi_k Y)^{d_k} .$$

A Point $P = (\xi : \eta : \zeta) \in \mathbb{P}^2_K(\bar{K})$ is a point of intersection of C and G if and only if $H(\xi, \eta) = 0$ and $\zeta = a'\xi + b'\eta$. The intersection points are therefore precisely the points

$$(\xi_1 : \eta_1 : a'\xi_1 + b'\eta_1) , \dots , (\xi_k : \eta_k : a'\xi_k + b'\eta_k) ,$$

and their multiplicities are by definition $d_1, \dots, d_k$, where $d_1 + \dots + d_k = d$. This proves the first part of the theorem.

For the second part we observe that we can write H as a product of d linear factors, $d-1$ of which have coefficients in L. Then the remaining factor also must have coefficients in L. ❑

This theorem is a special case of *Bézout's Theorem*, which says that two projective curves of degrees d_1 and d_2 that do not have a common component intersect in exactly $d_1 d_2$ points (counted with multiplicity; "point" means "$\bar{K}$-rational point" here). To formulate the theorem in this generality, one has to define the multiplicity of an intersection point of two arbitrary curves. This requires to dig a bit deeper into algebraic geometry than we can do here.

5. Smoothness

When doing analysis, one usually wants the objects one considers not to have corners or creases, but to be "smooth" (like for example manifolds). One uses differentiability properties to define this notion. We will proceed in a similar way for algebraic curves. We cannot take derivatives of functions in the sense of limits of quotients of differences (we don't have a suitable topology at our disposal), but we can simply use formal derivatives of polynomials by applying the usual differentiation rules. The partial derivatives showing up in the following definitions are to be understood in this way.

5.1. **Definition.**

DEF
smooth point
smooth curve
singularity

(1) An affine plane curve $C\colon f(x,y)=0$ is *smooth* at a point $P=(\xi,\eta)\in C(L)$, if at least one of the partial derivatives of f at the point P, $\frac{\partial f}{\partial x}(\xi,\eta)$ and $\frac{\partial f}{\partial y}(\xi,\eta)$, does not vanish.

(2) A projective plane curve $C\colon F(X,Y,Z)=0$ is *smooth* at a point $P=(\xi:\eta:\zeta)\in C(L)$, if

$$\Bigl(\frac{\partial F}{\partial X}(\xi,\eta,\zeta),\frac{\partial F}{\partial Y}(\xi,\eta,\zeta),\frac{\partial F}{\partial Z}(\xi,\eta,\zeta)\Bigr)\neq(0,0,0)\,.$$

(3) A point P at which C is not smooth is a *singular point* or a *singularity* of C. (Here C can be affine or projective.)

(4) An (affine or projective) curve C is *smooth*, if it is smooth at all points $P\in C(\bar{K})$. Otherwise C is *singular*. ♢

Note that in part (4) of the definition "all points" again means "all $\bar{K}$-rational points". A curve can be singular even though it is smooth at all K-rational points.

A point on an affine curve is smooth if and only if it is smooth on the projective closure (exercise).

5.2. **Examples.**

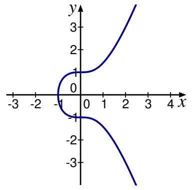
EX
smooth/
singular

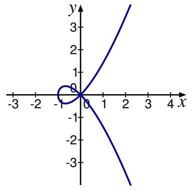
(1) Is the curve $Y^2Z-X^3-Z^3=0$ smooth? The points $(\xi:\eta:\zeta)$ at which the curve is not smooth must satisfy the following conditions.

$$-3\xi^2=2\eta\zeta=\eta^2-3\zeta^2=\eta^2\zeta-\xi^3-\zeta^3=0\,.$$

If we assume that $\operatorname{char}(K)\neq 2,3$, then this implies $\xi=\eta=\zeta=0$. Hence such a point cannot exist (since projective coordinates are not allowed to all vanish), and the curve is smooth.

In characteristic 2 the conditions are equivalent with $\xi=0$, $\eta=\zeta$; so the curve is singular at $(0:1:1)$. In characteristic 3 we obtain $\eta=0$ and $\xi+\zeta=0$; so the curve is singular at $(-1:0:1)$.





(2) The curve $y^2=x^3+x^2$ is not smooth at the point $P=(0,0)$, since both partial derivatives $3x^2+2x$ and $2y$ vanish there. In the picture "two branches cross" there; we have a so-called *simple double point* or *node*.

The sketches in the margin show the real points of the affine part of the curves. ♣

REM
Multiplicity of P on C

5.3. **Remark.** Let $C\colon F(X;Y,Z) = 0$ be a projective curve and let $P = (\xi : \eta : \zeta) \in C(K)$. It is not hard to show the following (exercise).

(1) C is smooth at P if and only if

$$i(C;P) := \min\{i(G,C;P) \mid G \text{ a line through } P\} = 1\,.$$

Otherwise, $i(C;P) \ge 2$. The number $i(C;P)$ is the *multiplicity* of P on C. Er have $i(C;P) \le d$, where d is the degree of the curve.

DEF
multiplicity of a point

(2) If C is smooth at P, then there exists exactly one line G through P such that $i(G,C;P) \ge 2$. This line is the *tangent* to C at P and is given by the equation

$$\frac{\partial F}{\partial X}(\xi,\eta,\zeta)\,X + \frac{\partial F}{\partial Y}(\xi,\eta,\zeta)\,Y + \frac{\partial F}{\partial Z}(\xi,\eta,\zeta)\,Z = 0\,.$$

DEF
tangent

If $i(G,C;P) = 3$, then P is an *inflection point* or *flex* of C; if $i(G,C;P) \ge 4$, then P is an *undulation point* or *hyperflex* of C. ♠

DEF
inflection point

6. Rational maps and morphisms

As always in mathematics, in algebraic geometry one considers not only objects (like algebraic curves), but also the relevant maps between them. We will now introduce these maps.

6.1. **Definition.** Let $C\colon F(X,Y,Z)=0$ and $D\colon G(X,Y,Z)=0$ be two irreducible projective plane curves over K.

DEF rational map morphism

(1) A *rational map* from C to D is an equivalence class of triples (R_1,R_2,R_3), where the $R_j \in K[X,Y,Z]$ are homogeneous of the same degree and not all divisible by F and such that F divides $G(R_1,R_2,R_3)$. Two triples (R_1,R_2,R_3) and (S_1,S_2,S_3) are equivalent, if $F \mid R_iS_j - R_jS_i$ for all i,j.

(2) Let ϕ be a rational map from C to D and let $P=(\xi:\eta:\zeta)\in C(L)$. ϕ is *regular* or *defined* at P, if ϕ has a representative (R_1,R_2,R_3) such that not all $R_j(\xi,\eta,\zeta)$ vanish. In this case

$$\phi_L(P) = (R_1(\xi,\eta,\zeta) : R_2(\xi,\eta,\zeta) : R_3(\xi,\eta,\zeta)) \in D(L)$$

is well-defined, and we obtain maps

$$\phi_L\colon \{P\in C(L) \mid \phi \text{ defined at } P\} \longrightarrow D(L)\,.$$

(3) A *morphism* from C to D is a rational map from C to D that is defined everywhere on C (i.e., on $C(\bar K)$).

(4) One can compose rational maps/morphisms in the obvious way. The equivalence class of (X,Y,Z) is the neutral element. The corresponding morphism is the identity morphism $\mathrm{id}_C\colon C\to C$.

(5) C and D are *birationally equivalent*, if there exist rational maps $\phi\colon C\to D$ and $\psi\colon D\to C$ such that $\phi\circ\psi=\mathrm{id}_D$ and $\psi\circ\phi=\mathrm{id}_C$. Then ϕ is a *birational* map. If ϕ and ψ are morphisms, then C and D are *isomorphic* and ϕ is an *isomorphism*. ♢

It is a fact that a rational map from a *smooth* curve to another curve is automatically a morphism. More precisely, a rational map from one curve C to another curve D is defined in every smooth point of C.

6.2. **Examples.**

EX morphisms

(1) Any two projective lines are isomorphic. For example, an isomorphism from $Z=0$ to $Z=aX+bY$ is given by

$$(X:Y:0)\mapsto(X:Y:aX+bY)\,.$$

(2) A morphism can be represented by constant polynomials. Such a constant morphism maps everything to a fixed (K-rational) point. One can show that every non-constant morphism between two irreducible projective curves is surjective; which means that $\phi_{\bar K}$ is surjective. (ϕ_L does not have to be surjective!)

(3) Here is a nontrivial example of a morphism. Let C be the "unit circle" $X^2+Y^2=Z^2$ over a field K with $\mathrm{char}(K)\neq 2$. Then $(X^2-Y^2, 2XY, Z^2)$ defines a morphism $\phi\colon C\longrightarrow C$: We have

$$(X^2-Y^2)^2+(2XY)^2-(Z^2)^2=(X^2+Y^2-Z^2)(X^2+Y^2+Z^2)\,,$$

so the essential condition is satisfied. The map is defined everywhere, since all three components vanish only for $X=Y=Z=0$, which does not correspond to a point on $\mathbb{P}^2_K$. (On the real unit circle $C(\mathbb{R})$ this morphism corresponds to doubling the angle from the positive x-axis.)

(4) Let C be the unit circle again and let $G\colon X = 0$ be the y-axis. Then the triples $(0, Y, Z + X)$ and $(0, Z - X, Y)$ define the same morphism $\phi\colon C \to G$, and $(Z^2 - Y^2, 2YZ, Z^2 + Y^2)$ defines a morphism $\phi'\colon G \to C$ that is inverse to ϕ (if $\operatorname{char}(K) \neq 2$): The unit circle (and therefore any circle that has K-rational points) is isomorphic to the y-axis (and therefore to any line)! (This even remains true for every irreducible *conic section*, i.e., a curve of degree 2, that has K-rational points.) ♣

7. Elliptic curves: definition

In this section we will introduce elliptic curves over an arbitrary base field.

What is an elliptic curve? The definition below appears to be somewhat ad-hoc, but is adequate for the purposes of this course, since we lack the necessary background from algebraic geometry for "better" definitions.

The "correct" definition is roughly as follows. An elliptic curve over K is an irreducible smooth projective curve over K of genus 1 together with a specified K-rational point on it. One can then show (using the Riemann-Roch Theorem) that an elliptic curve in this sense is always isomorphic to an elliptic curve in the sense of the definition below, with the isomorphism mapping the specified point to the point O.

We are elliptic curves called "elliptic"? There is a somewhat indirect connection with ellipses. If one wants to calculate the length of an arc on an ellipse, one arrives at an integral whose integrand has the form $R(x, \sqrt{P(x)})$, where R is a rational function in two variables and P is a polynomial of degree 3 or 4. Because of this, such integrals are also called elliptic integrals. If we write y for $P(x)$, then we have the relation $y^2 = P(x)$, which describes an (affine plane) curve that is birationally equivalent to an elliptic curve over any field K such that it has a K-rational point. One can view the integrand as a differential 1-form on this elliptic curve. Therefore one can say that elliptic curves are the curves on which elliptic integrals "live".

Elliptic curves themselves (as smooth irreducible curves of degree 3) are of course not ellipses (which are smooth irreducible curves of degree 2).

DEF elliptic curve

7.1. **Definition.** An *elliptic curve* over the field K is a smooth projective plane curve E of degree 3 over K that is given by an equation of the form

$$Y^2Z + a_1\,XYZ + a_3\,YZ^2 = X^3 + a_2\,X^2Z + a_4\,XZ^2 + a_6\,Z^3$$

with coefficients $a_1, a_2, a_3, a_4, a_6 \in K$.

For simplicity, we usually write the equation of the affine part

$$E\colon y^2 + a_1\,xy + a_3\,y = x^3 + a_2\,x^2 + a_4\,x + a_6\,. \qquad (7.1)$$

Such an equation is called a (long) *Weierstrass equation.* ◇

K. Weierstrass
1815–1897

The reason for the somewhat strange indexing of the coefficients will become clear soon. We already saw that such a curve is geometrically irreducible when $a_1 = a_2 = a_3 = 0$; the proof easily carries over to arbitrary curves of the form (7.1).

LEMMA point at infinity

7.2. **Lemma.** *Let E be a (not necessarily smooth) curve given by an equation of the form (7.1). Then E has exactly one point at infinity, which is $O = (0 : 1 : 0)$. This point O is K-rational, E is smooth at O, and the tangent line to E at O is the line at infinity $Z = 0$; it intersects E at O with multiplicity 3 (and hence O is an inflection point of E).*

Proof. To find the points at infinity, we set $Z = 0$ in the (projective) equation of the curve. This leaves $X^3 = 0$, hence the point $O = (0 : 1 : 0) \in E(K)$ is the only point at infinity, and the intersection multiplicity there with the line at infinity is 3. Since this multiplicity is ≥ 2 and E is smooth at O (see below), the line at infinity is the tangent line to E at O.

It remains to show that E is smooth at O. We show that the partial derivative with respect to Z does not vanish there. This derivative has the form Y^2 plus terms that contain X or Z, so evaluating it at $(0, 1, 0)$ gives 1. ❑

In many cases it is possible to simplify the equation of an elliptic curve.

7.3. **Lemma.** *Let E be an elliptic curve over K. If $\mathrm{char}(K) \neq 2$, then E is isomorphic (as an elliptic curve, see §8) to an elliptic curve of the form*

LEMMA
short
W. equation

$$E' \colon y^2 = x^3 + a_2'\,x^2 + a_4'\,x + a_6'\,.$$

If in addition $\mathrm{char}(K) \neq 3$, then one can obtain $a_2' = 0$. The resulting equation

$$y^2 = x^3 + ax + b$$

is a short Weierstrass equation.

Proof. The isomorphism from E to E' is given (in projective coordinates) by (this is "completing the square" on the left hand side)

$$(X : Y : Z) \longmapsto (X : Y + \frac{a_1}{2}\,X + \frac{a_3}{2}\,Z : Z)\,.$$

("Isomorphism *of elliptic curves*" means that the point $(0 : 1 : 0)$ of E is mapped to the point $(0 : 1 : 0)$ of E'.) The new coefficients then are

$$a_2' = a_2 + \tfrac{1}{4}a_1^2\,, \quad a_4' = a_4 + \tfrac{1}{2}a_1a_3\,, \quad a_6' = a_6 + \tfrac{1}{4}a_3^2\,.$$

If $\mathrm{char}(K) \neq 3$, then we can apply a further transformation of the form $(x, y) \mapsto (x + \frac{1}{3}a_2', y)$ to remove the coefficient a_2' ("completing the cube" on the right hand side). ❑

Now there is the obvious question when a (long or short) Weierstrass equation really defines an elliptic curve. Put differently, how can we determine whether the curve defined by the equation is smooth or not?

To answer this question we introduce a number of further quantities that depend on the coefficients. The notation is standard.

7.4. **Definition.** Let $E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ be a long Weierstrass equation. We set

DEF
$b_2, b_4, b_6, b_8,$
c_4, c_6, Δ, j

$$b_2 = a_1^2 + 4\,a_2\,, \quad b_4 = a_1a_3 + 2\,a_4\,, \quad b_6 = a_3^2 + 4\,a_6\,,$$
$$b_8 = a_1^2a_6 - a_1a_3a_4 + 4\,a_2a_6 + a_2a_3^2 - a_4^2$$
$$c_4 = b_2^2 - 24\,b_4\,, \quad c_6 = -b_2^3 + 36\,b_2b_4 - 216\,b_6$$
$$\Delta = -b_2^2b_8 - 8\,b_4^3 - 27\,b_6^2 + 9\,b_2b_4b_6\,, \quad j = c_4^3/\Delta$$

Then

$$4\,b_8 = b_2b_6 - b_4^2 \quad \text{and} \quad 1728\,\Delta = c_4^3 - c_6^2\,.$$

The quantities c_4 and c_6 are frequently called the *invariants* of the curve; Δ is the *discriminant* and j is the *j-invariant* of the curve. ♢

Note that the simplified equations (when $\mathrm{char}(K) \neq 2$, resp., $\mathrm{char}(K) \neq 2, 3$) can be written after an additional scaling of the variables ($(x, y) \mapsto (4x, 8y)$ and $(x, y) \mapsto (36x, 216y)$, respectively) in the form

$$y^2 = x^3 + b_2\,x^2 + 8b_4\,x + 16b_6$$

and

$$y^2 = x^3 - 27c_4\,x - 54c_6\,.$$

7.5. **Lemma.** *A Weierstrass equation of the form (7.1) defines an elliptic (i.e., a smooth) curve if and only if the discriminant Δ does not vanish.*

LEMMA
discriminant
and
smoothness

Proof. For simplicity, we restrict to the case $\operatorname{char}(K) \neq 2, 3$. The other cases can be dealt with in a similar way (but are more involved).

Assuming this, we can transform the original equation into a short Weierstrass equation $E\colon y^2 = x^3 + a\,x + b$; one can check that Δ changes by multiplication with the twelfth power of an invertible element (compare §8). Since we have an isomorphism between the two curves, either both or none of them are smooth. We then have

$$\Delta = -16(4\,a^3 + 27\,b^2)\,.$$

We have already seen that E is smooth at the point at infinity. We can therefore restrict to the affine part. An affine point (ξ, η) is singular on E if and only if the following three equations hold.

$$3\,\xi^2 + a = 0\,, \quad 2\,\eta = 0\,, \quad \eta^2 = \xi^3 + a\,\xi + b\,.$$

The assumption on the characteristic of K implies that this is equivalent to

$$\eta = 0\,, \quad \xi^2 = -\tfrac{1}{3}a\,, \quad \xi^3 + a\,\xi + b = 0\,.$$

Substituting the second equation in the third gives (if $a \neq 0$)

$$\xi = -\frac{3b}{2a}\,.$$

Hence the system has a solution if and only if

$$\left(\frac{3b}{2a}\right)^2 = -\frac{a}{3}\,,$$

which is equivalent to $\Delta = 0$.

When $a = 0$, the condition simplifies to $b = 0$, which is in this case again equivalent to $\Delta = 0$. ❑

This implies that when E is an elliptic curve over K, its j-invariant $j(E) = c_4^3/\Delta$ is a well-defined element of K.

7.6. **Examples.**

EX
discriminant

(1) The curve $y^2 = x^3$ has $\Delta = 0$, so it is not an elliptic curve. In fact, $(0, 0)$ is a singular point.

(2) The curve $y^2 = x^3 + x^2$ also has $\Delta = 0$ and a singularity at $(0, 0)$.

(3) The curve $y^2 = x^3 + x$ has $\Delta = -2^6$, so it is an elliptic curve if $\operatorname{char}(K) \neq 2$. Its j-invariant is $12^3 = 1728$.

(4) The curve $y^2 = x^3 + 1$ has $\Delta = -2^4 \cdot 3^3$, so it is an elliptic curve if $\operatorname{char}(K) \neq 2, 3$. Its j-invariant is 0. ♣

8. Isomorphisms of elliptic curves

In the last section we already referred to the notion of isomorphism of elliptic curves. We will now define it.

8.1. **Definition.** Let E and E' be two elliptic curves over K. A morphism $\phi\colon E \to E'$ is an *isomorphism of elliptic curves*, if ϕ has the form

$$(X : Y : Z) \longmapsto (u^2X + rZ : u^3Y + su^2X + tZ : Z)$$

with $r, s, t \in K$, $u \in K^\times$. ♢

DEF
isomorphism
of ell. curves

It is easy to see that a morphism of this form is indeed an isomorphism of curves; see below.

8.2. **Lemma.** *If E (resp., E') in Definition 8.1 is given by a Weierstrass equation with coefficients a_j (resp., a'_j), then*

$$\begin{aligned}
u\,a_1 &= a'_1 + 2s \\
u^2\,a_2 &= a'_2 - s\,a'_1 + 3r - s^2 \\
u^3\,a_3 &= a'_3 + r\,a'_1 + 2t \\
u^4\,a_4 &= a'_4 - s\,a'_3 + 2r\,a'_2 - (t + rs)\,a'_1 + 3r^2 - 2st \\
u^6\,a_6 &= a'_6 + r\,a'_4 - t\,a'_3 + r^2\,a'_2 - rt\,a'_1 + r^3 - t^2\,.
\end{aligned}$$

(This explains the indexing of the coefficients!) In addition, we have

$$u^4\,c_4 = c'_4\,, \quad u^6\,c_6 = c'_6\,, \quad u^{12}\,\Delta = \Delta' \quad \textit{and} \quad j = j'\,.$$

LEMMA
transformation
of the
coefficients

Proof. Computation. (Substitute into the equation of E'; the result must be a constant multiple of the equation of E; compare coefficients.) ❑

8.3. **Lemma.** *Let E and E' be elliptic curves over K. Then every isomorphism of elliptic curves $\phi\colon E \to E'$ is also an isomorphism of projective plane curves, and $\phi(O) = O$.*

An isomorphism of projective plane curves $\phi\colon E \to E'$ that is given by linear polynomials and maps O to O is an isomorphism of elliptic curves.

LEMMA
characterization of
isomorphisms

Proof. For the first claim, one checks that the inverse morphism is given by

$$\psi\colon (X : Y : Z) \longmapsto (u^{-2}(X - rZ) : u^{-3}(Y - sX + (sr - t)Z) : Z)\,.$$

Furthermore, $\phi(O) = (0 : u^3 : 0) = (0 : 1 : 0)$.

For the second claim, we can assume that ϕ has the form

$$(X : Y : Z) \longmapsto (\alpha_1X + \alpha_2Y + \alpha_3Z : \beta_1X + \beta_2Y + \beta_3Z : \gamma_1X + \gamma_2Y + \gamma_3Z)\,.$$

Since the line $Z = 0$ at infinity is the unique line that meets E and also E' at O with multiplicity 3 and since $\phi(O) = O$, ϕ must map this line to itself. This means that $\gamma_1 = \gamma_2 = 0$. That O is fixed by ϕ then means that $\alpha_2 = 0$. We can then set $\gamma_3 = 1$ without loss of generality (if $\gamma_3 = 0$, the morphism would be constant), and we see that the isomorphism has the required form, with the possible exception of the relation between α_1 and β_2. This relation follows by comparing coefficients after substituting into the Weierstrass equation, which gives $\alpha_1^3 = \beta_2^2$ and therefore th existence of some $u \in K$ such that $\alpha_1 = u^2$ and $\beta_2 = u^3$. Finally, u cannot vanish, since otherwise ϕ would be constant. ❑

8.4. **Remark.** The deeper algebraic-geometric reason for the form of the isomorphisms is that the rational function x $(= X/Z)$ has a pole of order 2 at O and is regular everywhere else, and all such functions have the form $ux + r$ with $r \in K$, $u \in K^\times$. Similarly, the rational function y $(= Y/Z)$ has a pole of order 3 at O and is regular everywhere else, and all such functions have the form $uy + sx + t$ with $s, t \in K$, $u \in K^\times$. Since O is supposed to be fixed, the orders of the poles there are preserved, which leads to the form of the isomorphism. ♠

REM
background

We see that the j-invariant $j(E)$ is preserved by isomorphisms (hence the name). This leads to the question whether the converse is also true: are two elliptic curves with the same j-invariant necessarily isomorphic? The following theorem shows that the answer if essentially "yes".

8.5. **Theorem.** *Let E and E' be two elliptic curves over K.*

THM
isomorphism
and
j-invariant

(1) *Assume* $\operatorname{char}(K) \neq 2, 3$. *$E$ and E' are isomorphic over K if and only if there exists some $u \in K^\times$ such that $c_4(E') = u^4\, c_4(E)$ and $c_6(E') = u^6\, c_6(E)$.*

(2) *If $j(E) = j(E')$, then E and E' are isomorphic over $\bar{K}$.*

(3) *For every $j \in K$ there is an elliptic curve E over K such that $j(E) = j$.*

Proof. For simplicity, we assume $\operatorname{char}(K) \neq 2, 3$ in all parts of the theorem.

(1) The given curves are by Lemma 7.3 and the remark preceding Lemma 7.5 isomorphic to the curves

$$\tilde{E}\colon y^2 = x^3 - 27c_4(E)\,x - 54c_6(E) \quad \text{and} \quad \tilde{E}'\colon y^2 = x^3 - 27c_4(E')\,x - 54c_6(E')\,.$$

It follows that E and E' isomorphic if and only if $\tilde{E}$ and $\tilde{E}'$ are isomorphic.

"‘$\Rightarrow$"’: This follows from Lemma 8.2.

"‘$\Leftarrow$"’: If $c_4(E') = u^4 c_4(E)$ and $c_6(E') = u^6 c_6(E)$, then Lemma 8.2 implies that $\tilde{E}$ and $\tilde{E}'$ are isomorphic via $(x, y) \mapsto (u^2\, x, u^3\, y)$.

(2) If $j(E) = j(E') = j$, then either $c_4(E) = c_4(E') = 0 = j$ or $c_6(E) = c_6(E') = 0$, $j = 1728$, or $j \neq 0, 1728$ and $c_6(E)^2/c_4(E)^3 = c_6(E')^2/c_4(E')^3 \neq 0$. In all three cases there is some $u \in \bar{K}^\times$ such that $c_4(E') = u^4\, c_4(E)$ and $c_6(E') = u^6\, c_6(E)$. The curves are then isomorphic by par (1).

(3) One checks that the cases $j = 0$ and $j = 12^3 = 1728$ are taken care of by the twp curves

$$y^2 = x^3 + 1 \qquad \text{and} \qquad y^2 = x^3 + x\,.$$

In the other cases, one can use the curve

$$y^2 = x^3 - \frac{27}{4}\,\frac{j}{j - 1728}\,x - \frac{27}{4}\,\frac{j}{j - 1728}\,.$$

(One obtains this curve by setting $a = b$ in the short Weierstrass equation $y^2 = x^3 + ax + b$.) ❏

So if K is algebraically closed, then the elliptic curves over K are classified up to isomorphism by the j-invariant. If K is not algebraically closed, then there can be several non-isomorphic elliptic curves with the same j-invariant.

THM
ell. curves
with $j(E) = j$

8.6. **Theorem.** *Assume that* $\mathrm{char}(K) \neq 2, 3$, *let* $j \in K$ *and* $E\colon y^2 = x^3 + a\,x + b$ *an elliptic curve over* K *with* $j(E) = j$.

(1) *In the case* $j \neq 0, 1728$, *the* K*-isomorphism classes of elliptic curves* E' *with* $j(E') = j$ *are classified by* $K^\times/(K^\times)^2$. *If* $d \in K^\times$ *represents such a class, then the associated elliptic curve is given by*

$$y^2 = x^3 + d^2\,a\,x + d^3\,b\,.$$

This curve is the quadratic twist *with* d *of* E.

DEF
quadratic
twist

(2) *If* $j = 0$, *then* $a = 0$. *The* K*-isomorphism classes with* $j = 0$ *are classified by* $K^\times/(K^\times)^6$*; the elliptic curve associated to* $d \in K^\times$ *is*

$$y^2 = x^3 + d\,b\,.$$

(3) *If* $j = 1728$, *then* $b = 0$. *The* K*-isomorphism classes with* $j = 1728$ *are classified by* $K^\times/(K^\times)^4$*; the elliptic curve associated to* $d \in K^\times$ *is*

$$y^2 = x^3 + d\,a\,x\,.$$

Proof. We have that $j = 0 \iff c_4 = 0 \iff a = 0$ and $j = 1728 \iff c_6 = 0 \iff b = 0$.

(1) The j-invariant of a short Weierstrass equation is a fractional linear function of a^3/b^2; in this case we have $a, b \neq 0$. Therefore $E'\colon y^2 = x^3 + a'x + b'$ has the same j-invariant as E if and only if $a'^3/b'^2 = a^3/b^2$, which is equivalent to $a' = d^2\,a$ and $b' = d^3\,b$ for some $d \in K^\times$. By Theorem 8.5, the two curves are already isomorphic over K if and only if d is a square.

(2) and (3) are shown in an analogous way. ❑

9. Group structure

Now we want to show that an elliptic curve carries a (geometrically defined) group structure.

9.1. Theorem. *Let E be an elliptic curve over K and let $L \supset K$ be a field extension. The following specifications determine a structure as abelian group on $E(L)$.* **THM** Group structure

(1) *The point $O \in E(L)$ is the zero element.*

(2) *If is G a line that meets E in the points P, Q, R (a point occurs as many times as given by its multiplicity as an intersection point), then $P + Q + R = O$.*

Somewhat more concretely, this means

(1) The point $-P$ is the third intersection point of the line through O and P with E.

(2) The point $P + Q$ is the third intersection point of the line through O and R with E, where R is the third intersection point of the line through P and Q with E.

Of course, we have to count all points with the correct multiplicity. For example, when P and Q are the same point, we have to consider the tangent line to E in $P = Q$ (instead of the line through P and Q), since this is the only line that intersects E in this point with multiplicity at least 2.

Formulas for addition.

To make this even more concrete, let E be given by the equation

$$E\colon\, y^2 + a_1\,xy + a_3\,y = x^3 + a_2\,x^2 + a_4\,x + a_6\,,$$

and let P and Q be the affine points (ξ, η) and (ξ', η'). The line through P and O is given by

$$x = \xi$$

and the third intersection point is

$$-P = (\xi, -\eta - a_1\,\xi - a_3)\,.$$

When $\xi \neq \xi'$, the line through P and Q is given by

$$y = \lambda\,x + \mu$$

with

$$\lambda = \frac{\eta' - \eta}{\xi' - \xi} \quad \text{and} \quad \mu = \eta - \lambda\,\xi = \frac{\xi'\eta - \xi\eta'}{\xi' - \xi}\,.$$

When $\xi = \xi'$ and $\eta + \eta' \neq -a_1\,\xi - a_3$ (the latter means that $Q \neq -P$), then we have $\eta = \eta'$ and

$$\lambda = \frac{3\,\xi^2 + 2a_2\,\xi + a_4 - a_1\,\eta}{2\,\eta + a_1\,\xi + a_3} \quad \text{and} \quad \mu = \eta - \lambda\,\xi = \frac{-\xi^3 + a_4\,\xi + 2a_6 - a_3\,\eta}{2\,\eta + a_1\,\xi + a_3}\,.$$

To see this, we can either determine the equation of the tangent to E in P, or we figure out that one can rewrite the quotient of differences in the case $\xi \neq \xi'$ using the equation of E as follows.

$$\begin{aligned}\frac{\eta' - \eta}{\xi' - \xi} &= \frac{(\eta'^2 + a_1\,\xi'\eta' + a_3\,\eta') - (\eta^2 + a_1\,\xi\eta + a_3\,\eta) - a_1(\xi' - \xi)\eta'}{(\xi' - \xi)(\eta' + \eta + a_1\,\xi + a_3)} \\ &= \frac{(\xi' - \xi)(\xi'^2 + \xi'\xi + \xi^2 + a_2(\xi' + \xi) + a_4 - a_1\,\eta')}{(\xi' - \xi)(\eta' + \eta + a_1\,\xi + a_3)} \\ &= \frac{\xi'^2 + \xi'\xi + \xi^2 + a_2(\xi' + \xi) + a_4 - a_1\,\eta'}{\eta' + \eta + a_1\,\xi + a_3}\end{aligned}$$

We can then replace ξ' and η' by ξ and η, respectively.

The coordinates of the third intersection point $R = (\xi'', \eta'')$ of this line with E then satisfy

$$\xi + \xi' + \xi'' = \lambda^2 + a_1\,\lambda - a_2\,, \quad \text{and hence} \quad \xi'' = \lambda^2 + a_1\,\lambda - a_2 - \xi - \xi'\,.$$

This can be seen by substituting $y = \lambda\,x + \mu$ into the equation defining E:

$$x^3 - (\lambda^2 + a_1\lambda - a_2)x^2 - (2\lambda\mu + a_1\mu + a_3\lambda - a_4)x - (\mu^2 + a_3\mu - a_6) = 0$$

ξ, ξ', ξ'' are the three solutions of this equations, hence their sum equals the negative of the coefficient of x^2.

Finally (using $\eta'' = \lambda\,\xi'' + \mu$) we obtain

$$P + Q = -R = (\xi'', -(\lambda + a_1)\xi'' - \mu - a_3)\,.$$

In a simplified form (for short Weierstrass equations) we have seen these formulas already in the introductory section.

Before we prove the theorem, we formulate a lemma that we will need to show that our addition is associative.

9.2. **Lemma.** *Let G_i and G'_j (for $i, j \in \{1, 2, 3\}$) be pairwise distinct lines in the projective plane such that the nine intersection points P_{ij} of G_i and G'_j are pairwise distinct. Let further C be a plane projective curve of degree 3 that contains the eight points P_{ij} with $(i, j) \neq (3, 3)$. Then C also contains the ninth point P_{33}.*

LEMMA
Lines and
curves of
degree 3

Proof. Let G_i and G'_j be given by $L_i(X, Y, Z) = 0$ resp. $L'_j(X, Y, Z) = 0$ with linear polynomials L_i, L'_j.

There are 10 monomials of degree 3 in three variables. The condition $P_{ij} \in C$ results in a homogeneous linear equation for the ten coefficients of C. The space of homogeneous polynomials of degree 3 that vanish in the eight given points therefore has at least dimension 2. In any case, this space contains the polynomials $L = L_1L_2L_3$ and $L' = L'_1L'_2L'_3$, which are linearly independent. We show that the dimension is in fact exactly 2, i.e., the space is spanned by L and L'.

To this end we assume that the dimension is at least 3. Then we can prescribe two additional points to lie on C. We choose a point P on G_1 that is not one of the intersection points of G_1 with the lines G'_j and a point Q that is not on any of the lines G_i. (Since $\#\mathbb{P}^2(\mathbb{F}_2) = 7 < 9$, the assumptions imply that K has at least three elements. Then the lines have $\#K + 1 \geq 4$ points, so that we can pick a suitable P. For Q we can take a point on G'_1 distinct from P_{11}, P_{21} and P_{31}.) Let $C\colon F(X, Y, Z) = 0$ be a curve of degree 3 that contains the eight given points and in addition P and Q. Since G_1 meets this curve in the four points P_{1j} $(j = 1, 2, 3)$ and P, Bézout's Theorem 4.3 implies that L_1 divides F: $F = L_1 F'$ with a homogeneous polynomial F' of degree 2. The curve of degree 2 defined by F' meets the line G_2 in the three points P_{2j} $(j = 1, 2, 3)$, therefore L_2 must divide F': $F' = L_2 F''$. Finally, the line defined by F'' shares with G_3 the two points P_{31} and P_{32}; so the two lines must be the same. It follows that $F = cL$ with some constant c. But this contradicts $Q \in C$, since Q does not lie on any of the lines G_i. This contradiction shows that the dimension is in fact only 2.

Let now $C\colon F = 0$ be a curve of degree 3 through the eight points. We have just shown that we must then have $F = cL + c'L'$ with constants c and c'. Since the right hand side vanishes at the point P_{33}, the same must be true of the left hand side, which means that $P_{33} \in C$. ❑

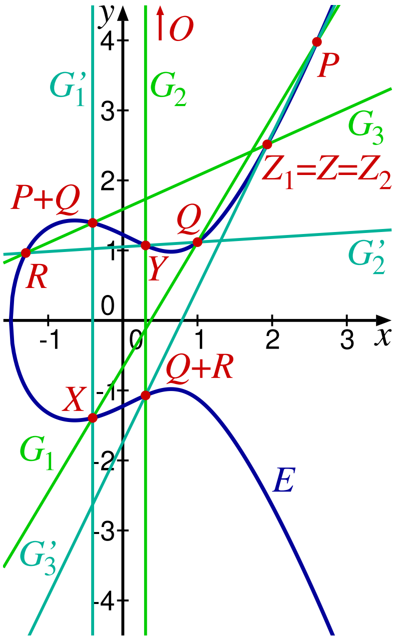
With this preparation we can approach the proof of Theorem 9.1.

Proof of Theorem 9.1. According to the second part of Theorem 4.3, the addition map on $E(L)$ is well-defined, since the line through two L-rational points (or the tangent to E in an L-rational point) is defined over L, hence the third intersection point is also in $E(L)$. The point O is by definition the zero element, and we have already seen that every point P has an inverse $-P$. Commutativity is also clear, since the construction of the sum $P + Q$ is symmetric in P and Q. It remains to show the associative law

$$(P + Q) + R = P + (Q + R)\,.$$

We consider the following objects.

G_1 is the line through P and Q;
X is its third intersection point with E.
G_1' is the line through O and X;
its third intersection point with E is $P+Q$.
G_2' is the line through Q and R;
Y is its third intersection point with E.
G_2 is the line through O and Y;
its third intersection point with E is $Q+R$.
G_3 is the line through $P+Q$ and R;
Z_1 is its third intersection point with E.
G_3' is the line through $Q+R$ and P;
Z_2 is its third intersection point with E.
Z is the intersection point of G_3 and G_3'.



We first assume that the nine points O, P, Q, R, X, Y, $P+Q$, $Q+R$ and Z are pairwise distinct. Since

$$Z_1 = -((P+Q)+R) \qquad \text{and} \qquad Z_2 = -(P+(Q+R))\,,$$

it is enough to show that $Z_1 = Z = Z_2$.

We now want to apply Lemma 9.2 to our lines G_i and G_j'. These lines are all distinct, since otherwise we would have at least four points in the intersection of E with one of the lines (five or six of the nine points lie on the union of two of the lines; one of them could be the point Z, for which we do not know yet that it is on E), which would imply by Theorem 4.3 that the line is contained in E.

However, E is irreducible and therefore cannot contain a line as a component. So the lemma is applicable. We have the following identifications.

$$\begin{array}{lll} P_{11} = X\,, & P_{12} = Q\,, & P_{13} = P\,, \\ P_{21} = O\,, & P_{22} = Y\,, & P_{23} = Q + R\,, \\ P_{31} = P + Q\,, & P_{32} = R\,, & P_{33} = Z\,. \end{array}$$

Furthermore, E is a curve of degree 3 through the first eight points, so the lemma implies that $Z \in E$. Hence Z is the third intersection point both of G_3 and of G'_3 with E; in particular, $Z_1 = Z = Z_2$.

This proves the associative law in the “generic” case. The cases in which points coincide can either be treated one by one, or one uses a kind of “continuity argument”—the morphism

$$E \times E \times E \ni (P, Q, R) \mapsto \big((P + Q) + R\big) - \big(P + (Q + R)\big) \in E$$

takes the constant value O an an “open and dense” subset and must therefore be constant. Of course, we neither have defined the product on the left hand side, nor what a morphism is in this context, nor what the *Zariski topology* is that comes into play. Therefore we modify this argument so that it works with a morphism $E \to E$.

We first note that the associative law holds trivially when $P = O$ or $R = O$ or $P = R$ (in the last case we use commutativity). If $P, R \neq O$ and $P \neq R$, then there are only finitely many points Q such that the nine intersection points in the argument above are not pairwise distinct (exercise; for this it is useful to first show that for a non-constant morphism $\psi\colon E \to E$ the equation $\psi(S) = T$ for a given point $T \in E$ has only finitely many solutions). We consider the morphism

$$\phi_{P,R}\colon E \longrightarrow E\,, \quad Q \longmapsto ((P + Q) + R) + (-(P + (Q + R)))\,.$$

The equivalence $P + (-Q) = O \iff P = Q$ is easily seen. This implies that $\phi_{P,R}(Q) = O$ for all except finitely many $Q \in E(\bar{K})$; but then $\phi_{P,R}$ has to be constant $= O$. This means that the equality $(P + Q) + R = P + (Q + R)$ is valid for *all* Q. $\square$

One can prove the associative law also by using the explicit addition formulas that we have derived above. One then has to show that the expressions for the coordinates of $(P + Q) + R$ and of $P + (Q + R)$ agree modulo the equation of E evaluated in P, Q and R. If one does this by hand, it is very cumbersome; with a computer algebra system this is possible without problems, however.

REM
Group structure intrinsically

9.3. **Remark.** One can also characterize the group structure in an “intrinsic” way, as follows. Let P, Q, R be points of E. The relation $P + Q = R$ holds precisely when there is a rational function ϕ on E that has simple zeros in P and Q, simple poles in R and O, and no other zeros or poles. (If some of the four points O, P, Q, R coincide, one has to combine the order of the zeros and poles accordingly.)

One implication is easy to see. Let $L_1(X, Y, Z) = 0$ be the equation of the line through P and Q, and let $L_2(X, Y, Z) = 0$ be the equation of the line through R and O. Then $\phi = L_1/L_2$ is a suitable function: the numerator vanishes in P, Q and $-R$, and the denominator vanishes in R, O and $-R$, so that we find the required zeros and poles (the zeros of the numerator and the denominator at $-R$ “cancel each other out”).

This characterization implies that every isomorphism of curves $E \to E'$ that maps O to O is also compatible with the group structure. For our explicit definition of isomorphisms of elliptic curves this claim follows from the fact that such an isomorphism is linear and therefore maps lines to lines. So triples of intersection points of the curve with a line are mapped to similar triples, and so the group structure is preserved.

One prerequisite for this characterization is however that one has to define the order of a zero or a pole of a rational function on a curve. See the following section. ♠

9.4. **Example.** We use the definition of the group structure to find out which points on an elliptic curve have order 2, respectively, 3. We assume again that the characteristic of the base field K is not 2 or 3; we can therefore work with a short Weierstrass equation

EX points of order 2, 3

$$E\colon y^2 = x^3 + ax + b\,.$$

We assume in addition that K is algebraically closed (or declare "point" to mean "$\bar{K}$-rational point").

A point $P \in E(\bar{K})$ has order 2, if $P \neq O$ and $2P = P + P = O$. The latter is equivalent to $P + P + O = O$; by definition, this means that P, P, O are the three intersection points of E with some line. Since this line contains O, it is vertical. Such a line has an equation of the form $x = \xi$ for some fixed ξ; it meets the affine part of E in the (in general) two points with x-coordinate ξ. These points have the form (ξ, η) and $(\xi, -\eta)$. In our case the two must be the same; this means $\eta = 0$ (indeed, the tangent to E in a point with vanishing y-coordinate is vertical). We see that there are exactly three points of order 2, to wit, the points $(\xi, 0)$ with $\xi^3 + a\xi + b = 0$.

A point $P \in E(\bar{K})$ has order 3, if $P \neq O$ and $3P = P + P + P = O$. This means by definition that there exists a line that meets E in P with multiplicity 3. This means in turn that P is an inflection point of E (and the line is the inflectional tangent). One can show that a smooth cubic curve has always exactly nine inflection points (if $\operatorname{char}(K) \neq 3$). One of these is O, so there are exactly eight points of order 3. ♣

In characteristic 2, there is no or exactly one point of order 2, depending on whether $a_1 = 0$ or $a_1 \neq 0$; in the second case it is the point (ξ, η) with $\xi = -a_3/a_1$ and $\eta = \sqrt{\xi^3 + a_2\xi^2 + a_4\xi + a_6}$ (note that there are unique square roots in characteristic 2).

In characteristic 3 the situation is similar: there is either no or there are exactly two points of order 3. The formulas are a bit more involved; the condition for the existence of points of order 3 is $a_1^2 + a_2 \neq 0$.

10. Local rings and divisors

To make precise what we mean be the order of vanishing of a rational function at a point, we introduce the local ring of a curve at a point and show that it is a discrete valuation ring when the point is smooth.

10.1. **Definition.** Let C be an irreducible curve over K and let $P \in C(K)$. Then the ring

$$\mathcal{O}_{C,P} = \{\phi \in K(C) \mid \phi \text{ is regular at } P\}$$

is called the *local ring* of C at the point P. We write

$$\mathfrak{m}_P = \{\phi \in \mathcal{O}_{C,P} \mid \phi(P) = 0\}$$

for its unique maximal ideal. ♢

DEF
local ring
at a point

Recall that a ring R is *local* if it has a unique maximal ideal. This is equivalent to the statement that the complement of the unit group $R^\times$ is an ideal M (which is then the unique maximal ideal): any proper ideal I of R must satisfy $I \cap R^\times = \emptyset$, so $I \subset R \setminus R^\times = M$. On the other hand, assume that M is the unique maximal ideal. Take any $r \in R \setminus R^\times$. Then r is contained in a maximal ideal, so $r \in M$, showing that $R \setminus R^\times \subset M$. The reverse inclusion is obvious.

In our case, we see that every $\phi \in \mathcal{O}_{C,P} \setminus \mathfrak{m}_P$ is regular at P with $\phi(P) \neq 0$, which implies that ϕ^{-1} is also regular at P, so $\phi^{-1} \in \mathcal{O}_{C,P}$, whence $\phi \in \mathcal{O}_{C,P}^\times$.

10.2. **Definition.** Let R be a domain (i.e., a commutative ring without zero divisors). A *discrete valuation* on R is a surjective map $v\colon R \to \mathbb{Z}_{\geq 0} \cup \{\infty\}$ with the following properties, which hold for all $r, r' \in R$:

DEF
DVR

(1) $v(r) = \infty \iff r = 0$.

(2) $v(rr') = v(r) + v(r')$.

(3) $v(r + r') \geq \min\{v(r), v(r')\}$.

A domain R together with a discrete valuation v on it such that every $r \in R$ with $v(r) = 0$ is in $R^\times$ and such that the ideal $\{r \in R \mid v(r) > 0\}$ is principal is a *discrete valuation ring* or short *DVR*. ♢

10.3. **Lemma.** *Let (R, v) be a DVR. Then R is a local ring with unique maximal ideal $M = \{r \in R \mid v(r) > 0\}$ and unit group $R^\times = \{r \in R \mid v(r) = 0\}$. Also, R is a principal ideal domain (PID) with only one prime (up to associates): let $t \in R$ be an element such that $v(t) = 1$ (such a t is called a* uniformizer *of R); then every $r \in R \setminus \{0\}$ can be written uniquely in the form $r = ut^n$ with a unit $u \in R^\times$ and $n \in \mathbb{Z}_{\geq 0}$.*

LEMMA
Properties
of DVRs
DEF
uniformizer

Conversely, every PID with exactly one prime ideal $\neq 0$ (i.e., a local PID that is not a field) is a DVR.

Proof. Exercise. ❑

If R is a DVR with field of fractions K, then v extends to a valuation on K in a unique way by setting $v(r/s) = v(r) - v(s)$. Then the extended v is a surjective map $v\colon K \to \mathbb{Z} \cup \{\infty\}$ satisfying the conditions in Definition 10.2. We call (K, v) a *discretely valued field*.

DEF
discretely
valued field

EX
DVR $\mathbb{Z}_{(p)}$

10.4. **Example.** Let p be a prime number and set

$$\mathbb{Z}_{(p)} = \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z}, p \nmid b \right\}.$$

Then $\mathbb{Z}_{(p)}$ is a DVR with the discrete valuation given by the p-adic valuation v_p ($v_p(a/b) = v_p(a) = \max\{n \mid p^n \mid a\}$). Its field of fractions is $\mathbb{Q}$, which becomes a discretely valued field with the valuation v_p. ♣

LEMMA
local ring at P is DVR

10.5. **Lemma.** *Let C be an irreducible curve over K and let $P \in C(K)$. If C is smooth at P, then the local ring $\mathcal{O}_{C,P}$ is a DVR with field of fractions $K(C)$.*

The converse is also true: if $\mathcal{O}_{C,P}$ is a DVR, then C is smooth at P (exercise).

Proof. After a suitable coordinate transformation, we can assume that P is the affine point $(0,0)$. Since C is smooth at P, the affine equation $f(x,y) = 0$ defining it has linear terms; without loss of generality, we can assume that

$$f(x,y) = \alpha x + \beta y + (\text{terms of degree} \geq 2)$$

with $\beta \neq 0$. Then $f(x,y) = y(\beta + y h_1(y)) + x h_2(x,y)$ with polynomials $h_1 \in K[y]$, $h_2 \in K[x,y]$.

In the following, we will abuse notation and write x and y for the corresponding functions in $K(C)$ (i.e., the images of x and y under $K[x,y] \to K[C] \to K(C)$); in particular, $f(x,y) = 0$.

We now show that $y/x \in \mathcal{O}_{C,P}$. Since

$$0 = f(x,y) = y(\beta + y h_1(y)) + x h_2(x,y)\,,$$

we have

$$\frac{y}{x} = -\frac{h_2(x,y)}{\beta + y h_1(y)}\,,$$

where the numerator is in $\mathcal{O}_{C,P}$ (all polynomials in x and y are regular at P) and the denominator does not vanish at P (since $\beta \neq 0$), so it is a unit in $\mathcal{O}_{C,P}$. This shows that the quotient is in $\mathcal{O}_{C,P}$ as claimed.

Next, we show that $\mathfrak{m}_P = \langle x \rangle_{\mathcal{O}_{C,P}}$. Since clearly $x \in \mathfrak{m}_P$, the inclusion "$\supset$" is obvious. For the other inclusion, take $\phi \in \mathfrak{m}_P$. We can then write $\phi = \phi_1(x,y)/\phi_2(x,y)$ with polynomials ϕ_1, ϕ_2 such that $\phi_1(0,0) = 0$ and $\phi_2(0,0) \neq 0$. The latter implies that $\phi_2(x,y) \in \mathcal{O}_{C,P}^\times$. The former means that we can write $\phi_1(x,y) = x\psi_1(x,y) + y\psi_2(x,y)$ with some polynomials ψ_1 and ψ_2. Then

$$\phi = \phi_2(x,y)^{-1}\Big(\psi_1(x,y) + \frac{y}{x}\psi_2(x,y)\Big) \cdot x \in \langle x \rangle\,,$$

since $\phi_2(x,y)^{-1} \in \mathcal{O}_{C,P}$ and $y/x \in \mathcal{O}_{C,P}$.

We now invoke the Krull intersection theorem from commutative algebra, which tells us that $\bigcap_{n \geq 0} \mathfrak{m}_P^n = \{0\}$ (since $\mathcal{O}_{C,P}$ is a noetherian local ring; the noetherian property (every ideal is finitely generated) is inherited from the polynomial ring $K[x,y]$). This implies that for each $0 \neq \phi \in \mathcal{O}_{C,P}$ there is a unique $n = n(\phi) \geq 0$ such that $\phi \in \langle x^n \rangle \setminus \langle x^{n+1} \rangle$; this implies that $\phi = u x^n$ with $u \in \mathcal{O}_{C,P}^\times$.

Let $I \subset \mathcal{O}_{C,P}$ be a nonzero ideal. Let n be the minimum over all $n(\phi)$ with $0 \neq \phi \in I$; this is a well-defined non-negative integer. I claim that $I = \langle x^n \rangle$. For "$\subset$", observe that every $0 \neq \phi \in I$ is (by the definition of n) an element of

$\langle x^{n(\phi)} \rangle \subset \langle x^n \rangle$. For "$\supset$", let $0 \neq \phi \in I$ be such that $n(\phi) = n$. Then $\phi = ux^n$ with a unit u, so $x^n = u^{-1}\phi \in I$.

This shows that every ideal of $\mathcal{O}_{C,P}$ is principal, so $\mathcal{O}_{C,P}$ is a local PID. Since $\mathcal{O}_{C,P}$ is not a field (there are nonzero non-units, for example x), it must be a DVR.

Finally, the field of fractions of $\mathcal{O}_{C,P}$ is contained in $K(C)$ and it contains x and y, so it contains $K(x, y) = K(C)$. (This argument does not use that C is smooth at P.) ❑

DEF
P-adic valuation
uniformizer at P
order of vanishing
pole

10.6. **Definition.** In the situation of Lemma 10.5, we write v_P for the extension of the discrete valuation of $\mathcal{O}_{C,P}$ to its field of fractions $K(C)$ and call it the *P-adic valuation* of $K(C)$. Any element $t \in K(C)$ such that $v_P(t) = 1$ is called a *uniformizer at P*.

A function $\phi \in K(C)^\times$ is regular at P if and only if $v_P(\phi) \geq 0$. In this case, we say that $v_P(\phi)$ is the *order of vanishing* of ϕ at P. Otherwise, we say that ϕ has a *pole of order $-v_P(\phi)$* at P. ♢

In this sense, ϕ has a simple zero at P when $v_P(\phi) = 1$, and a simple pole at P when $v_P(\phi) = -1$.

In the following, we assume that K is algebraically closed.

DEF
divisor group
divisor
degree
effective
support

10.7. **Definition.** Let C be a smooth, projective and irreducible curve over K. The free abelian group with basis the set of K-points on C is called the *divisor group* of C, written Div_C. Its elements, which are formal integral linear combinations of points in $C(K)$, are *divisors* on C. We will usually write a divisor D as

$$D = \sum_{P \in C(K)} n_P \cdot (P)\,,$$

where $n_P \in \mathbb{Z}$ and $n_P = 0$ for all but finitely many points P. We also denote n_P by $v_P(D)$. The *degree* of such a divisor is $\deg(D) = \sum_P n_P$; this defines a homomorphism $\deg\colon \operatorname{Div}_C \to \mathbb{Z}$. The set of divisors of degree zero forms a subgroup Div_C^0 of Div_C. We write $D \geq D'$ if $v_P(D) \geq v_P(D')$ for all points P. A divisor D such that $D \geq 0$ is *effective*. The *support* of D is the set

$$\operatorname{supp}(D) = \{P \in C(K) \mid v_P(D) \neq 0\}$$

of points occurring in D with a nonzero coefficient. ♢

Now if $\phi \in K(C)^\times$ is a nonzero rational function on C, then it is easy to see (considering a representative quotient of polynomials) that ϕ has only finitely many zeros and poles on C. The following definition therefore makes sense.

DEF
principal divisor
Picard group
linear equivalence

10.8. **Definition.** Let $\phi \in K(C)^\times$. We set

$$\operatorname{div}(\phi) = \sum_{P \in C(K)} v_P(\phi) \cdot (P)$$

and call this the *divisor of ϕ*. A divisor of this form is *principal*. We write Princ_C for the subgroup of principal divisors. The quotient group

$$\operatorname{Pic}_C = \operatorname{Div}_C / \operatorname{Princ}_C$$

is the *Picard group* of C. Two divisors D, D' are *linearly equivalent* if $D - D'$ is principal; we write $D \sim D'$. We usually write $[D]$ for the linear equivalence class of a divisor $D \in \operatorname{Div}_C$, i.e., for the image of D in Pic_C. ♢

Note that by the properties of valuations, the map

$$\operatorname{div}\colon K(C)^\times \longrightarrow \operatorname{Div}_C$$

is a group homomorphism, so its image Princ_C is a subgroup.

10.9. **Example.** Let $E\colon y^2 = x^3 + ax + b$ be an elliptic curve (with $\operatorname{char}(K) \neq 2, 3$). The divisor of the function x is

EX divisors of functions

$$\operatorname{div}(x) = (0, \sqrt{b}) + (0, -\sqrt{b}) - 2 \cdot (O)\,.$$

The divisor of y is

$$\operatorname{div}(y) = \sum_{\xi\colon \xi^3+a\xi+b=0} (\xi, 0) - 3 \cdot (O)\,.$$

Note that any polynomial in x and y is regular on the affine part of E, so only the point O at infinity can occur with a negative coefficient in the divisor of such a function.

We check that $v_O(x) = -2$ and $v_O(y) = -3$. Following the proof of Lemma 10.5, we consider the affine patch where $Y \neq 0$, which has equation

$$z' = x'^3 + ax'z'^2 + bz'^3\,.$$

This shows that $x' = X/Y = x/y$ is a uniformizer at O, so $v_O(x/y) = 1$. Also, $z'/x' \in \mathcal{O}_{E,O}$ and z'/x' vanishes at O, so

$$\frac{z'}{x'^3} = 1 + a\Bigl(\frac{z'}{x'}\Bigr)^2 + b\Bigl(\frac{z'}{x'}\Bigr)^3$$

is a unit in $\mathcal{O}_{E,O}$, hence $v_O(z') = 3v_O(x') = 3$. Then $x = X/Z = x'/z'$ has valuation $v_O(x') - v_O(z') = 1 - 3 = -2$ and $y = Y/Z = 1/z'$ has valuation $-v_O(z') = -3$. ♣

The following observation is useful. When v is a discrete valuation on a field K and $a, b \in K$ satisfy $v(a) < v(b)$, then $v(a + b) = v(a)$. (We have "$\ge$" by the defining properties of a valuation. We also have $v(-1) = 0$ (since $2v(-1) = v((-1)^2) = v(1) = 0$), so $v(a) = v((a + b) - b) \ge \min\{v(a + b), v(b)\} = v(a + b)$, since $v(a) < v(b)$.)

10.10. **Lemma.** *Let C be an irreducible curve over K and $P \in C(K)$ such that C is smooth at P. Let $G_1\colon L_1(X, Y, Z) = 0$ and $G_2\colon L_2(X, Y, Z) = 0$ be two lines that are not contained in C. Then*

LEMMA order of vanishing and intersection multiplicity

$$v_P\Bigl(\frac{L_1(X, Y, Z)}{L_2(X, Y, Z)}\Bigr) = i(G_1, C; P) - i(G_2, C; P)\,.$$

Proof. We can as usual assume that P is the affine point $(0, 0)$. Let $f(x, y) = 0$ be the affine equation of C. We can make another coordinate change if necessary to arrange that the lines are not vertical, so that $L_j(x, y, 1) = \alpha_j x - y + \beta_j$ for $j = 1, 2$ (after scaling the equations if necessary), and in addition

$$f(x, y) = y + x^d h_1(x) + xyh_2(x, y) + y^2 h_3(x, y)$$

with $d \ge 2$ (so $y = 0$ is the tangent line at P) and $h_1(0) \neq 0$ (if h_1 is the zero polynomial, then C contains the line $y = 0$, hence must be this line; we leave this case as an exercise); then $x = X/Z$ is a uniformizer at P and

$$v_P(y) = v_P\bigl(y(1 + xh_2(x, y) + yh_3(x, y))\bigr) = v_P(-x^d h_1(x)) = d\,.$$

We have that

$$v_P\Big(\frac{L_1(X,Y,Z)}{L_2(X,Y,Z)}\Big) = v_P(\alpha_1 x - y + \beta_1) - v_P(\alpha_2 x - y + \beta_2)\,,$$

so it is sufficient to show that $v_P(\alpha x - y + \beta) = i(G, C; P)$ where $G\colon \alpha X - Y + \beta Z = 0$. If $\beta \neq 0$, then both sides are zero. If $\beta = 0$, then $i(G, C; P)$ is the order of vanishing of

$$f(x, \alpha x) = \alpha x + x^d h_1(x) + \alpha x^2 h_2(x, \alpha x) + \alpha^2 x^2 h_3(x, \alpha x)$$

at $x = 0$, which is 1 if $\alpha \neq 0$ and d if $\alpha = 0$. On the other hand, $v_P(\alpha x - y) = 1$ if $\alpha \neq 0$ by the observation above, and $v_P(-y) = v_P(y) = d$. ❑

The divisors given in Example 10.9 already hint at the following fact.

10.11. **Lemma.** *Let C be a smooth, projective, irreducible curve over K and let $\phi \in K(C)^\times$. Then* $\deg \operatorname{div}(\phi) = 0$.

LEMMA
$\mathrm{Princ}_C \subset \mathrm{Div}^0_C$

Proof. We prove this for elliptic curves. One can prove it in a similar way for arbitrary curves, using some non-constant morphism $C \to \mathbb{P}^1$ (and more theory).

For simplicity, we assume that our curve is given by a short Weierstrass equation $y^2 = f(x) := x^3 + ax + b$. The automorphism $\iota\colon P \mapsto -P$ (which is just $(x, y) \mapsto (x, -y)$) acts on $K(C)^\times$ and on Div_C via its action on the points, and clearly $v_{-P}(\phi \circ \iota) = v_P(\phi)$. This implies that $\deg \operatorname{div}(\phi \circ \iota) = \deg \operatorname{div}(\phi)$ and therefore $\deg \operatorname{div}(\phi \cdot (\phi \circ \iota)) = 2 \deg \operatorname{div}(\phi)$. ϕ is represented by a function on $\mathbb{P}^2_g$ of the form $h_1(x) + h_2(x)y$ with $h_1, h_2 \in K(x)$ (this is because $y^2 = f(x)$); then $\phi \circ \iota = h_1(x) - h_2(x)y$ and

$$\phi \cdot (\phi \circ \iota) = h_1(x)^2 - h_2(x)^2 y^2 = h_1(x)^2 - h_2(x)^2 f(x) \in K(x)$$

is a function of x alone. We can write this projectively as a quotient of homogeneous polynomials in X and Z of the same degree d, which we can factor (recall that K is assumed to be algebraically closed) into d linear factors each, so that $\phi \cdot (\phi \circ \iota)$ is a product of d quotients Q_j of linear forms. By Lemma 10.10 and the special case of Bézout's Theorem 4.3, we see that $\deg \operatorname{div}(Q_j) = \deg C - \deg C = 0$ for each of these factors, which implies that $2 \deg \operatorname{div}(\phi) = \deg \operatorname{div}(\phi \cdot (\phi \circ \iota)) = 0$ as well. ❑

This means that Princ_C is contained in Div^0_C, so that deg descends to a homomorphism $\mathrm{Pic}_C \to \mathbb{Z}$. We denote its kernel by Pic^0_C.

DEF
Pic^0_C

Let now E be an elliptic curve over K. Then we can define a map

$$\alpha\colon E(K) \longrightarrow \mathrm{Pic}^0_E\,, \quad P \longmapsto [(P) - (O)]\,.$$

We will show that α is a group isomorphism.

10.12. **Lemma.** *This map α is a group homomorphism.*

LEMMA
canonical map is homom.

Proof. Clearly $\alpha(O) = 0$. It then suffices to show that $P_1 + P_2 + P_3 = O$ implies $\alpha(P_1) + \alpha(P_2) + \alpha(P_3) = 0$, i.e., that $(P_1) + (P_2) + (P_3) - 3 \cdot (O)$ is a principal divisor. Now since $P_1 + P_2 + P_3 = O$, there is a line G such that P_1, P_2 and P_3 are the intersection points of G and E (counted with multiplicity); let $L(X, Y, Z) = 0$ be an equation for G and define $\phi = L(X, Y, Z)/Z \in K(E)^\times$. By Lemma 10.10, $v_Q(\phi) = i(G, C; Q) - i(G_\infty, C; Q)$ for each point $Q \in E(K)$, where G_∞ is the line at infinity. This implies that $\operatorname{div}(\phi) = (P_1) + (P_2) + (P_3) - 3 \cdot (O)$, so this divisor is indeed principal. ❑

10.13. **Lemma.** *The map α is injective.*

LEMMA canonical map is injective

Proof. Since α is a group homomorphism, it suffices to show that $\alpha(P) = 0$ implies that $P = O$. $\alpha(P) = 0$ means that $(P) - (O)$ is a principal divisor, so there is some $\phi \in K(E)^\times$ such that $\operatorname{div}(\phi) = (P) - (O)$. Since ϕ is regular away from O, $\phi \in K[E]$ and so it is a polynomial in x and y. Since we can express y^2 as a polynomial in x plus y times a polynomial in x via the equation of E, we can write ϕ (uniquely) in the form $\phi = h_1(x) + h_2(x)y$ with polynomials h_1 and h_2. Let d_j be the degree of h_j, for $j = 1, 2$ (set $d_j = -\infty$ if h_j is the zero polynomial). Then $v_O(h_1(x)) = -2d_1$ and $v_O(h_2(x)y) = -2d_2 - 3$. Since one of these is even (if finite) and the other is odd (if finite) and not both can be infinite, these valuations are distinct, and so

$$v_P(\phi) = \min\{-2d_1, -2d_2 - 3\} \in \{0, -2, -3, \ldots\}$$

cannot be equal to -1. So the only possibility is that $(P) = (O)$ (and ϕ is constant). ❑

10.14. **Lemma.** *The map $\alpha\colon E(K) \to \operatorname{Pic}^0_E$ is surjective.*

LEMMA canonical map is surjective

Proof. Let $[D] \in \operatorname{Pic}^0_E$ with $D \in \operatorname{Div}^0_E$, so $D = \sum_P n_P \cdot (P)$ with $\sum_P n_P = 0$. Then $D = \sum_P n_P((P) - (O))$, so

$$[D] = \Big[\sum_P n_P\big((P) - (O)\big)\Big] = \sum_P n_P\big[(P) - (O)\big] = \sum_P n_P \alpha(P) = \alpha\Big(\sum_P n_P \cdot P\Big),$$

where in the last step we use that α is a group homomorphism and the linear combination in the last expression is taken in the group $E(K)$. This shows that $[D]$ is in the image of α. ❑

10.15. **Theorem.** *Let E be an elliptic curve over K and let $D = \sum_P n_P \cdot (P)$ be a divisor on E. Then D is principal if and only if $\deg D = 0$ and $\sum_P n_P \cdot P = O$ in $E(K)$.*

THM Abel-Jacobi theorem for elliptic curves

Proof. By the preceding three lemmas, the map

$$\alpha\colon E(K) \longrightarrow \operatorname{Pic}^0_E, \quad P \longmapsto \big[(P) - (O)\big]$$

is a group isomorphism. Since a principal divisor has degree zero, $\deg D = 0$ is a necessary condition. So assume that $\deg D = 0$. Then

$$[D] = \Big[\sum_P n_P\big((P) - (O)\big)\Big] = \sum_P n_P \alpha(P) = \alpha(\sum_P n_P \cdot P),$$

and since α is injective, this vanishes if and only if $\sum_P n_P \cdot P = O$. ❑

This also provides the intrinsic definition of the group law:

$$P + Q = R \iff P + Q = R + O \iff (P) + (Q) - (R) - (O) \sim 0\,.$$

This definition can be used whenever E is a (smooth projective irreducible) curve with a specified point O such that the map α is a bijection. (We effectively transfer the group structure from Pic^0_E to $E(K)$ via α.) One can show that this is equivalent to the curve having genus 1.

11. Isogenies and endomorphisms

The relevant maps between elliptic curves are morphisms that respect the group structure. Before we introduce them, we need some more results on the relation between rational maps and function fields.

11.1. **Theorem.** *Let C and D be two irreducible (projective) curves over K. Then there is a bijection $\phi \mapsto \phi^*$ between the set of all non-constant rational maps $\phi\colon C \to D$ over K and the set of all K-linear homomorphisms $\phi^*\colon K(D) \to K(C)$ between the function fields. Here $K(C)$ is a finite field extension of $\phi^*\bigl(K(D)\bigr)$.*

THM rational maps and function fields

If E is another curve and $\psi\colon D \to E$ is another non-constant rational map, then we have $(\psi \circ \phi)^* = \phi^* \circ \psi^*$.

Proof. (Sketch) It is a bit simpler to formulate this using affine coordinates. To do this, we assume that neither C nor D is the line at infinity $Z = 0$ (otherwise one needs to adapt the argument slightly). Let C' and D' be the affine parts of C and D; then $K(C') = K(C)$ and $K(D') = K(D)$. We write x and y for the affine coordinate functions on C' and u and v for those on D'. Then $K(C) = K(x, y)$ and $K(D) = K(u, v)$.

A rational map $\phi\colon C \to D$ is given in affine coordinates by two rational functions $r(x, y), s(x, y) \in K(x, y) = K(C)$ such that $\bigl(r(x, y), s(x, y)\bigr)$ is a $K(C)$-rational point on D' (note that r and s satisfy the affine equation of D'):

$$\phi\colon (x, y) \longmapsto \bigl(r(x, y), s(x, y)\bigr)\,.$$

ϕ is non-constant when r and s are not both constant (i.e., in $\bar{K} \cap K(C)$). The associated homomorphism of the function fields is then given by

$$\phi^*\colon K(D) \ni f \longmapsto f \circ \phi \in K(C)\,.$$

Expressed in coordinates, this means

$$\phi^*(u) = r(x, y)\,, \quad \phi^*(v) = s(x, y)\,.$$

A homomorphism of fields is always injective (since the only proper ideal of a field is the zero ideal). The field extension $K(C)/\phi^*\bigl(K(D)\bigr)$ is finite, since x and y are algebraic over $\phi^*\bigl(K(D)\bigr) = K\bigl(r(x, y), s(x, y)\bigr)$.

Conversely, if a K-linear homomorphism $\psi\colon K(D) \to K(C)$ is given, then we set $r(x, y) = \psi(u)$, $s(x, y) = \psi(v)$. Let $g(u, v) = 0$ be the equation of D', then we have

$$g\bigl(r(x, y), s(x, y)\bigr) = g\bigl(\psi(u), \psi(v)\bigr) = \psi\bigl(g(u, v)\bigr) = \psi(0) = 0\,,$$

and so

$$\phi\colon C \longrightarrow D\,, \qquad (x, y) \longmapsto \bigl(r(x, y), s(x, y)\bigr)$$

is a (non-constant) rational map such that $\psi = \phi^*$.

Finally, we have for $f \in K(E)$

$$(\psi \circ \phi)^*(f) = f \circ (\psi \circ \phi) = (f \circ \psi) \circ \phi = \phi^*(\psi^*(f)) = (\phi^* \circ \psi^*)(f)\,. \qquad ❑$$

11.2. **Definition.** In the situation of Theorem 11.1, the degree of the field extension $\phi^*\big(K(D)\big) \subset K(C)$ is the *degree* of ϕ, written $\deg \phi$.

DEF
degree of ϕ

Let $\phi^*\big(K(D)\big) \subset L \subset K(C)$ the maximal intermediate field that is separable over $\phi^*\big(K(D)\big)$. Then the degree $[L : \phi^*\big(K(D)\big)]$ is the *separable degree* of ϕ, $\deg_s \phi$, and the degree $[K(C) : L]$ is the *inseparable degree* of ϕ, $\deg_i \phi$. We then clearly have $\deg \phi = (\deg_s \phi)(\deg_i \phi)$.

separable

ϕ is *separable* if $L = K(C)$ (this is always the Case when $\operatorname{char}(K) = 0$), otherwise ϕ is *inseparable*. ϕ is *purely inseparable* if $L = \phi^*\big(K(D)\big)$. ♢

11.3. **Corollary.** *Two irreducible curves C and D over K are birationally equivalent over K if and only if their function fields $K(C)$ and $K(D)$ are isomorphic as field extensions of K.*

COR
birational
equivalence

Proof. "‘$\Rightarrow$"’: Let $\phi\colon C \to D$ be a birational map with inverse $\psi = \phi^{-1}$. By Theorem 11.1, we then have $\psi^* \circ \phi^* = (\phi \circ \psi)^* = \mathrm{id}_{K(D)}$ and $\phi^* \circ \psi^* = (\psi \circ \phi)^* = \mathrm{id}_{K(C)}$, hence ϕ^* is a K-linear isomorphism of $K(D)$ and $K(C)$.

"‘$\Leftarrow$"’: Let $\alpha\colon K(D) \to K(C)$ be a K-linear isomorphism with inverse isomorphism $\beta = \alpha^{-1}$. By Theorem 11.1, there is a rational map $\phi\colon C \to D$ with $\alpha = \phi^*$ and a rational map $\psi\colon D \to C$ with $\beta = \psi^*$. Then

$$\begin{aligned}(\psi \circ \phi)^* &= \phi^* \circ \psi^* = \alpha \circ \beta = \mathrm{id}_{K(C)} \qquad \text{and}\\ (\phi \circ \psi)^* &= \psi^* \circ \phi^* = \beta \circ \alpha = \mathrm{id}_{K(D)}\,;\end{aligned}$$

this shows that ϕ and ψ are inverse rational maps, and so C and D are birationally equivalent. ❑

11.4. **Corollary.** *Let C_1, C_2 and C_3 be irreducible curves over K and let*

$$\phi_1\colon C_1 \to C_2 \qquad \textit{and} \qquad \phi_2\colon C_2 \to C_3$$

be non-constant rational maps. Then

$$\begin{aligned}\deg(\phi_2 \circ \phi_1) &= (\deg \phi_2)(\deg \phi_1)\,,\\ \deg_s(\phi_2 \circ \phi_1) &= (\deg_s \phi_2)(\deg_s \phi_1) \qquad \textit{and}\\ \deg_i(\phi_2 \circ \phi_1) &= (\deg_i \phi_2)(\deg_i \phi_1)\,.\end{aligned}$$

COR
degree is
multiplicative

Proof. The first equality follows from the multiplicativity of the degree in the tower of fields $K(C_3) \hookrightarrow K(C_2) \hookrightarrow K(C_1)$.

For the second equality, let L be the maximal separable intermediate field of the extension $K(C_3) \hookrightarrow K(C_1)$, let L' be the maximal separable intermediate field of $K(C_3) \hookrightarrow K(C_2)$, and let L'' be the maximal separable intermediate field of $K(C_2) \hookrightarrow K(C_1)$. Then we have $L'' = K(C_2) \cdot L$ and $[L : L'] = [L'' : K(C_2)]$. This implies

$$\begin{aligned}\deg_s(\phi_2 \circ \phi_1) = [L : K(C_3)] &= [L : L'] \cdot [L' : K(C_3)]\\ &= [L'' : K(C_2)] \cdot [L' : K(C_3)] = (\deg_s \phi_1)(\deg_s \phi_2)\,.\end{aligned}$$

(See also the Wikipedia entry on separable extensions.) The third equality follows from the first two. ❑

Now we can introduce the relevant maps between elliptic curves. It turns out (like for isomorphisms) that it suffices to require the minimal assumption $\phi(O) = O$.

11.5. **Definition.** Let E and E' be elliptic curves over K. An *isogeny* from E to E' is a morphism $\phi\colon E \to E'$ such that $\phi(O) = O$. The curves E and E' are *isogenous*, if there exists a non-constant isogeny $E \to E'$. ♢

DEF
isogeny

Such an isogeny is therefore either constant: $\phi(P) = O$ for all $P \in E$, or else surjective (as a map $\phi_{\bar{K}}\colon E(\bar{K}) \to E'(\bar{K})$).

In the literature the constant map $P \mapsto O$ is not always considered to be an isogeny.

⚠

The most important property of isogenies is that they automatically respect the group structures on E and E'.

11.6. **Theorem.** *Let $\phi\colon E \to E'$ by an isogeny. Then*

$$\phi_L(P+Q) = \phi_L(P) + \phi_L(Q) \quad \textit{for all } P, Q \in E(L),$$

i.e., ϕ is a group homomorphism.

THM
isogeny is
homomorphism

Proof. See for example [Si1, Thm. III.4.8]. Here is a sketch. We have seen that the sum $P+Q$ is characterized by the fact that there exists a rational function f on E with divisor $(P) + (Q) - (P+Q) - (O)$. If $\phi\colon E \to E'$ is a non-constant morphism with $\phi(O) = O$, then we can consider the norm of f, $N(f) \in K(E')$, with respect to the field extension given by ϕ^*. This function $N(f)$ has divisor $(\phi(P)) + (\phi(Q)) - (\phi(P+Q)) - (\phi(O))$; since $\phi(O) = O$, this shows that $\phi(P+Q) = \phi(P) + \phi(Q)$. ❑

The most important examples of isogenies are the *multiplication maps*. Let $m \in \mathbb{Z}$ and let E be an elliptic curve. Then

DEF
multiplication map

$$[m] = [m]_E \colon E \ni P \longmapsto [m](P) = m \cdot P \in E$$

defines an isogeny (where $m \cdot P$ is the mth multiple of P as an element of an abelian group ($= \mathbb{Z}$-module)).

11.7. **Lemma.** *Let E be an elliptic curve and $m \in \mathbb{Z} \setminus \{0\}$. Then $[m_E]$ is not constant.*

LEMMA
$[m]$ is
non-constant

Proof. (See also [Si1, Prop. III.4.2.(a)]). As we have seen earlier in Example 9.4, there are at most four points $P \in E(\bar{K})$ such that $2P = O$, so $[2]$ cannot be constant. If $\operatorname{char}(K) \neq 2$, then there are exactly three such points $P \neq O$, and then $[m](P) = mP = P \neq O$ for every odd m, and so $[m]$ is not constant. Since the composition of two non-constant isogenies is again non-constant, this suffices to show the claim.

If $\operatorname{char}(K) = 2$, then again by Example 9.4, there are exactly eight points $P \in E(\bar{K}) \setminus \{O\}$ such that $3P = O$. Similarly as above, this shows that $[3]$ is not constant, and, picking one of these points, we have that $[m](P) = \pm P \neq O$ for every m that is not divisible by 3. ❑

11.8. **Definition.** Isogenies $E \to E$ (like for example the multiplication maps) are also called *endomorphisms* of E; they form a ring $\mathrm{End}_K(E)$ (which is a subring of the endomorphism ring of the abelian group $E(\bar{K})$) with point-wise defined sum, $(\phi + \psi)(P) = \phi(P) + \psi(P)$, and with composition as product, $\phi \cdot \psi = \phi \circ \psi$. ♢

DEF
endomorphism

Like for every non-constant rational map between curves, we have the degree $\deg \phi$, the separable degree $\deg_s \phi$ and the inseparable degree $\deg_i \phi$ for every non-constant isogeny $\phi\colon E \to E'$. For completeness, one sets $\deg 0 = \deg_s 0 = \deg_i 0 = 0$ (where the first three 0s are the constant isogeny $P \mapsto O$). We then have

$$\deg(\psi \circ \phi) = (\deg \psi)(\deg \phi)\,, \qquad \deg \phi \geq 0 \qquad \text{and} \qquad \deg \phi = 0 \iff \phi = 0\,.$$

11.9. **Theorem.** *The endomorphism ring* $\mathrm{End}_K(E)$ *is a ring of characteristic* 0 *without zero divisors.*

THM
endomorphism
ring

Proof. Let $\phi, \psi \in \mathrm{End}_K(E)$ with $\phi \cdot \psi = 0$. This implies $0 = \deg(\phi\psi) = \deg(\phi)\deg(\psi)$, hence $\deg(\phi) = 0$ or $\deg(\psi) = 0$ and then $\phi = 0$ or $\psi = 0$. This shows that $\mathrm{End}_K(E)$ has no zero divisors.

Furthermore, by Lemma 11.7 $[m] = 0$ (i.e., constant) only when $m = 0$; the homomorphism $\mathbb{Z} \ni m \mapsto [m] \in \mathrm{End}_K(E)$ is therefore injective. This means that the endomorphism ring has characteristic zero. ❑

In particular, we always have the embedding $\mathbb{Z} \ni m \mapsto [m] \in \mathrm{End}_K(E)$.

One can classify pretty exactly which rings can occur as an endomorphism ring of an elliptic curve. In characteristic zero, $\mathrm{End}_K(E) = \mathbb{Z}$ is most common. Over finite fields, however, the endomorphism ring is always larger, since one has in addition the *Frobenius endomorphism* $(x, y) \mapsto (x^q, y^q)$ (where q is the size of the base field). We will come back to this later in more detail.

11.10. **Remark.** The following statements are valid more generally for isogenies between possibly distinct elliptic curves:

REM
properties
of isogenies

$$\begin{aligned}
\forall \phi_1, \phi_2\colon E \to E', \psi\colon E' \to E''\colon &\quad \psi \circ (\phi_1 + \phi_2) = \psi \circ \phi_1 + \psi \circ \phi_2 \\
\forall \phi\colon E \to E', \psi_1, \psi_2\colon E' \to E''\colon &\quad (\psi_1 + \psi_2) \circ \phi = \psi_1 \circ \phi + \psi_2 \circ \phi \\
\forall \phi\colon E \to E', \psi\colon E' \to E''\colon &\quad \psi \circ \phi = 0 \iff (\psi = 0 \text{ or } \phi = 0)
\end{aligned}$$

Here the sum of two isogenies $E \to E'$ is again defined point-wise like in Definition 11.8. ♠

The following property is also very important.

11.11. **Theorem.** *Let* $\phi\colon E \to E'$ *be a non-constant isogeny of degree* m. *Then there exists a unique isogeny* $\hat{\phi}\colon E' \to E$, *the* dual isogeny *of* ϕ *such that* $\hat{\phi} \circ \phi = [m]_E$. *Then we also have* $\phi \circ \hat{\phi} = [m]_{E'}$. *Furthermore:*

THM
dual
isogeny

(1) *If* $\psi\colon E' \to E''$ *is another isogeny, then* $\widehat{\psi \circ \phi} = \hat{\phi} \circ \hat{\psi}$.

(2) *If* $\psi\colon E \to E'$ *is another Isogeny, then* $\widehat{\phi + \psi} = \hat{\phi} + \hat{\psi}$.

(3) $\hat{\hat{\phi}} = \phi$.

(4) $\deg(\hat{\phi}) = \deg(\phi)$.

(5) *For all* $m \in \mathbb{Z}$ *we have* $\widehat{[m]}_E = [m]_E$ *and* $\deg([m]_E) = m^2$.

We also set $\hat{0} = 0$*; then* (1)–(5) *are valid for arbitrary isogenies.*

Proof. See for example [Si1, Thms III.6.1 and 6.2]. We do not show the existence of $\hat{\phi}$ here. To show uniqueness, let $\psi, \psi' \colon E' \to E$ be isogenies with $\psi \circ \phi = \psi' \circ \phi = [m]$. This implies $(\psi - \psi') \circ \phi = \psi \circ \phi - \psi' \circ \phi = 0$, and since $\phi \neq 0$, we must have $\psi - \psi' = 0$, so $\psi = \psi'$.

From $\hat{\phi} \circ \phi = [m]_E$ we can also deduce that

$$(\phi \circ \hat{\phi}) \circ \phi = \phi \circ (\hat{\phi} \circ \phi) = \phi \circ [m]_E = [m]_{E'} \circ \phi \,,$$

which by a similar argument implies $\phi \circ \hat{\phi} = [m]_{E'}$.

We now show some of the properties claimed in the theorem.

(1) Let $m = \deg \phi$, $n = \deg \psi$; then $\deg(\psi \circ \phi) = nm$. The isogeny $\hat{\phi} \circ \hat{\psi}$ satisfies

$$(\hat{\phi} \circ \hat{\psi}) \circ (\psi \circ \phi) = \hat{\phi} \circ (\hat{\psi} \circ \psi) \circ \phi = \hat{\phi} \circ [n] \circ \phi = [n] \circ (\hat{\phi} \circ \phi) = [n] \circ [m] = [nm] \,,$$

so it must be equal to $\widehat{\psi \circ \phi}$ by uniqueness.

(5) One can show $\deg[m] = m^2$ by establishing an explicit recurrence relation for the functions $r_m(x)$ that give the x-coordinate of mP for a point $P = (x, y)$ (see Lemma 11.12 below). Since $[m] \circ [m] = [m^2] = \big[\deg[m]\big]$, we then also obtain $\widehat{[m]} = [m]$.

(4) We have $m^2 = \deg[m] = (\deg \hat{\phi})(\deg \phi) = (\deg \hat{\phi})m$; this implies $\deg \hat{\phi} = m = \deg \phi$.

(3) We have $\phi \circ \hat{\phi} = [m] = [\deg \hat{\phi}]$, hence $\phi = \hat{\hat{\phi}}$.

(2) We do not prove this here. ❑

LEMMA
form of isogenies

11.12. **Lemma.** *If* $\phi \colon E \to E'$ *is a non-constant isogeny between elliptic curves given by Weierstrass equations of the form* $y^2 = f(x)$*, then* ϕ *has the form*

$$(x, y) \longmapsto (r(x), s(x)y) \,,$$

where $r(x)$ *and* $s(x)$ *are quotients of polynomials in* x*. If* $r(x) = p(x)/q(x)$ *is given in lowest terms, then* $\deg \phi = \max\{\deg p, \deg q\}$*.*

Proof. Since y satisfies a quadratic equation over $K(x)$, we can write every rational map $E \to E'$ uniquely in the form

$$(x, y) \longmapsto (r_1(x) + r_2(x)y, s_1(x) + s_2(x)y) \,,$$

where r_1, r_2, s_1, s_2 are rational functions of x. Since ϕ is a homomorphism, we have $\phi(-P) = -\phi(P)$, i.e., $\phi(x, -y) = -\phi(x, y)$, which expands to

$$(r_1(x) - r_2(x)y, s_1(x) - s_2(x)y) = (r_1(x) + r_2(x)y, -s_1(x) - s_2(x)y) \,.$$

Therefore r_2 and s_1 must be zero.

Writing x' and y' for the affine coordinate functions on E', we furthermore have that $[K(E') : K(x')] = [K(x')(y') : K(x')] = 2$ and $[K(E) : K(x)] = 2$, and also

$$[K(x) : K(x')] = [K(x) : K\big(r(x)\big)] = \max\{\deg p, \deg q\} \,.$$

The multiplicativity of degrees in towers of field extensions then implies that

$$\begin{aligned}\deg\phi &= [K(E):K(E')] \\ &= \frac{[K(E):K(x)][K(x):K(x')]}{[K(E'):K(x')]} \\ &= [K(x):K(x')] = \max\{\deg p, \deg q\}\,. \end{aligned}$$ ❑

The statement that the x-coordinate of $\phi(x,y)$ has the form $r(x)$ holds more generally also for long Weierstrass equations. The same is true for the formula giving the degree of ϕ.

Now it is time for an example.

11.13. **Example.** Let K be a field with $\operatorname{char}(K) \neq 2$ and let

EX isogeny

$$E\colon y^2 = x^3 + a\,x^2 + b\,x$$

be an elliptic curve over K. (This means that $b \neq 0$ and $a^2 - 4b \neq 0$.) By Example 9.4, we know that the point $(0,0) \in E(K)$ has order 2. The equation

$$E'\colon y^2 = x^3 - 2a\,x^2 + (a^2 - 4b)\,x$$

also defines an elliptic curve over K, and we have the two dual isogenies

$$\begin{aligned}\phi\colon E \longrightarrow E', \quad (x,y) &\longmapsto \left(\frac{y^2}{x^2}, \frac{b-x^2}{x^2}\,y\right) = \left(\frac{x^2+ax+b}{x}, \frac{b-x^2}{x^2}\,y\right) \\ \hat{\phi}\colon E' \longrightarrow E, \quad (x,y) &\longmapsto \left(\frac{y^2}{4x^2}, \frac{a^2-4b-x^2}{8x^2}\,y\right)\end{aligned}$$

Lemma 11.12 shows that both have degree 2, and one checks by a computation that $\hat{\phi} \circ \phi = [2]_E$ and $\phi \circ \hat{\phi} = [2]_{E'}$, as has to be the case for the two isogenies to be dual to each other (exercise).

The kernel of ϕ clearly consists of the two points $O, (0,0) \in E(K)$; similarly, the kernel of $\hat{\phi}$ consists of the two points $O, (0,0) \in E'(K)$. That the size of the kernel equals the degree is not a coincidence; see Theorem 11.15 below. ♣

If we apply Theorem 11.11 on the dual isogeny to the endomorphism ring of E (whose elements are the isogenies $E \to E$), then we obtain the following result.

11.14. **Theorem.** *The map $\operatorname{End}_K(E) \to \operatorname{End}_K(E)$, $\phi \mapsto \hat{\phi}$, is an anti-involution of $\operatorname{End}_K(E)$ (i.e., an anti-automorphism that is its own inverse, where "anti" means that the order of factors in a product gets reversed). If we identify $\mathbb{Z}$ with its image in $\operatorname{End}_K(E)$, then we have*

THM dualizing in $\operatorname{End}_K(E)$

$$\phi + \hat{\phi} \in \mathbb{Z} \quad \textit{and} \qquad \phi\hat{\phi} = \deg(\phi)\,.$$

Furthermore, $\deg$ *defines a positive definite quadratic form on* $\operatorname{End}_K(E)$.

Proof. That dualizing is an anti-involution follows from parts (1)–(3) of Theorem 11.11. The first part of this theorem also shows that $\phi\hat{\phi} = \deg(\phi)$. In order to show $\phi + \hat{\phi} \in \mathbb{Z}$, we consider

$$\mathbb{Z} \ni \deg(1+\phi) = (1+\phi)\widehat{(1+\phi)} = (1+\phi)(1+\hat{\phi}) = 1 + \phi + \hat{\phi} + \deg(\phi)\,.$$

That $\deg$ is a quadratic form means that $\deg(n\phi) = n^2 \deg\phi$ for all $n \in \mathbb{Z}$ and that

$$(\phi, \psi) \longmapsto \deg(\phi+\psi) - \deg(\phi) - \deg(\psi)$$

is $\mathbb{Z}$-bilinear in ϕ and ψ. The first claim follows easily from Theorem 11.11: $\deg(n\phi) = (\deg n)(\deg \phi) = n^2 \deg \phi$. The second claim can be seen by re-writing the right hand side as $\phi\hat{\psi}+\psi\hat{\phi}$ and observing that $\phi \mapsto \hat{\phi}$ $\mathbb{Z}$-linear by Theorem 11.11 again. The quadratic form deg is positive definite, since $\deg(\phi) \ge 0$ for all ϕ and $\deg(\phi) = 0$ only when $\phi = 0$. ❑

It is the presence of this anti-involution inducing a positive definite quadratic form that makes a classification of all possible endomorphism rings possible (together with a further result that bounds the rank of $\mathrm{End}_K(E)$ as a $\mathbb{Z}$-module by 4).

We will need a result on the relation between the order of the kernel of an isogeny and its (separable) degree.

11.15. **Theorem.** *Let $\phi\colon E \to E'$ be a non-constant isogeny over K. Then we have for all $P \in E'(\bar{K})$ that*

$$\#\phi_{\bar{K}}^{-1}(P) = \deg_s(\phi)\,.$$

In particular, the kernel of $\phi_{\bar{K}}$ has order $\deg_s(\phi)$*.*

THM
kernel
and degree

Proof. Compare [Si1, Thm. III.4.10]. Roughly speaking, we obtain an algebraic equation of degree $\deg\phi$ for the x-coordinates of the preimages of $P = (\xi, \eta)$ (compare Lemma 11.12), which can be written in the form

$$f(x^{p^k}) = f_1(x^{p^k}) - \xi f_2(x^{p^k}) = 0\,,$$

where p^k is the inseparable degree of ϕ and $\deg_s \phi = \deg f = \max\{\deg f_1, \deg f_2\}$. (Here p is the characteristic of K. When $\mathrm{char}(K) = 0$, we set $p^k = \deg_i \phi = 1$.) Then $f(x)$ is a polynomial without multiple roots for all but finitely many ξ, so for almost all ξ, there are exactly $\deg_s \phi$ solutions of $f(x) = 0$ in $\bar{K}$, which lead to the same number of points $Q \in E(\bar{K})$ with $\phi_{\bar{K}}(Q) = P$. (Note that exactly one of the points with the given x-coordinate is mapped to P, the other one (if it exists) to $-P$.) Since the sets $\phi_{\bar{K}}^{-1}(P)$ for each P can be obtained from any one of them by translation (addition of a suitable point in $E(\bar{K})$), they all must have exactly $\deg_s \phi$ elements. ❑

In the case of characteristic $p \neq 0$ we will need to know when an isogeny is inseparable.

So let E be an elliptic curve over a field K of characteristic p and let $q = p^e$ be a power of p. If we replace every coefficient a_j in a Weierstrass equation of E by its qth power a_j^q, then we obtain an equation defining an elliptic curve $E^{(q)}$ over K (the discriminant of the new equation is the qth power of the discriminant of the original equation, hence non-zero). We also obtain an isogeny

$$\phi\colon E \longrightarrow E^{(q)}\,, \qquad (x, y) \longmapsto (x^q, y^q)\,.$$

If K is finite and q is a power of $\#K$, then $E^{(q)} = E$; when $q = \#K$, ϕ is called the *Frobenius endomorphism* of E.

DEF
Frobenius
endomorphism

LEMMA
inseparable
endomorphisms

11.16. **Lemma.** *Let $K = \mathbb{F}_q$ with $q = p^e$ and let E be an elliptic curve over K.*

(1) *Let $\phi\colon E \to E^{(p)}$, $(x, y) \mapsto (x^p, y^p)$. Then ϕ is purely inseparable: $\deg \phi = \deg_i \phi = p$.*

(2) *Let $\pi \in \operatorname{End}_K(E)$ be the Frobenius endomorphism and let $m, n \in \mathbb{Z}$. The endomorphism $m + n\pi$ is separable if and only if m is not divisible by p.*

Proof. Compare [Si1, Cor. III.5.5 and Prop. II.2.11]. Part (1) is clear, since we adjoin a pth root, which gives a purely inseparable extension. The statement on the degree follows from Lemma 11.12.

Since $\deg \phi = p$, we have $[p] = \hat{\phi}\phi$, which implies $\deg_i[p] \ge \deg_i \phi = p > 1$. We can decompose $\pi = \psi\phi$ (where $\psi\colon E^{(p)} \to E$, $(x, y) \mapsto (x^{p^{e-1}}, y^{p^{e-1}})$). If $m = pm'$ is divisible by p, then we have $m + n\pi = (m'\hat{\phi} + n\psi) \circ \phi$, which is inseparable, since ϕ is. That multiplication by m is separable when $p \nmid m$ follows from $p \nmid m^2 = \deg[m]$ and the fact that the inseparable degree is always a power of p. For the converse in (2) (which will be important for us), one needs the statement that the sum of a separable and an inseparable isogeny is separable. This can be seen (as is done in Silverman's book) by using the *invariant differential* of E. ❑

DEF
$E[\phi]$

11.17. **Definition.** If $\phi\colon E \to E'$ is an isogeny, then we write $E[\phi]$ for its kernel, i.e.,

$$E[\phi] = \ker \phi_{\bar{K}} = \{P \in E(\bar{K}) \mid \phi_{\bar{K}}(P) = O\}\,.$$

We write $E(K)[\phi]$ for the group of K-rational points in the kernel. If $\phi = [m]$ is a multiplication map, then we simply write $E[m]$ for its kernel (which is the group of points on E whose order divides m). ♢

DEF
τ_Q

For $Q \in E$ we define the translation map $\tau_Q\colon E \to E$ by $P \mapsto P + Q$.

LEMMA
$E[\phi]$ and
field ext.

11.18. **Lemma.** *Let $\phi\colon E \to E'$ be an isogeny. Then*

$$E[\phi] \longrightarrow \operatorname{Aut}\bigl(\bar{K}(E)/\phi^*(\bar{K}(E'))\bigr)\,, \qquad T \longmapsto (f \mapsto f \circ \tau_T)$$

is a group isomorphism.

If ϕ is separable and $f \in \bar{K}(E)$ such that $f \circ \tau_T = f$ for all $T \in E[\phi]$, then there is $f' \in \bar{K}(E')$ such that $f = f' \circ \phi$.

Proof. If $f \in \phi^*(\bar{K}(E'))$, so that $f = f' \circ \phi$ with some $f' \in \bar{K}(E')$, then $f = f \circ \tau_T$ for $T \in E[\phi]$, since

$$f(P + T) = f'(\phi(P + T)) = f'(\phi(P) + \phi(T)) = f'(\phi(P)) = f(P)\,.$$

Hence $f \mapsto f \circ \tau_T$ is indeed an automorphism of $\bar{K}(E)$ over $\phi^*(\bar{K}(E'))$, and the map in the statement is well-defined. Since $\tau_T \circ \tau_{T'} = \tau_{T+T'}$, the map is a group homomorphism. A general result on field extension says that

$$\#\operatorname{Aut}\bigl(\bar{K}(E)/\phi^*(\bar{K}(E'))\bigr) \le [\bar{K}(E) : \phi^*(\bar{K}(E'))]_s = \deg_s \phi = \#E[\phi]\,.$$

It is therefore sufficient to show that the map is injective. If it were not injective, then there would exist some $O \ne T \in E[\phi]$ such that $f \circ \tau_T = f$ for all $f \in \bar{K}(E)$. But this is wrong for $f = x$ for example, since this function has a pole only at O, whereas $f \circ \tau_T$ has a pole only at $-T$.

We then obtain the last claim as follows. Since ϕ is separable, the first claim implies that

$$\#\operatorname{Aut}\bigl(\bar{K}(E)/\phi^*(\bar{K}(E'))\bigr) = \deg\phi = [\bar{K}(E) : \phi^*(\bar{K}(E'))]\,,$$

hence $\phi^*(\bar{K}(E')) \subset \bar{K}(E)$ is a Galois extension. By the assumption on f and the first claim again, f is invariant under the action ofr $\operatorname{Gal}\bigl(\bar{K}(E)/\phi^*(\bar{K}(E'))\bigr)$; this implies that $f \in \phi^*(\bar{K}(E'))\bigr)$. ❑

12. Torsion and Weil pairing

In this section we will study the structure of the *m-torsion points* of an elliptic curve in some detail. These are the points P such that $m \cdot P = O$.

DEF m-torsion point

12.1. Theorem. *Let E be an elliptic curve over an algebraically closed field K and let $m \in \mathbb{Z}_{>0}$.*

THM Torsion

(1) *If* $\operatorname{char}(K)$ *does not divide m (e.g.,* $\operatorname{char}(K) = 0$*), then*

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/m\mathbb{Z}\,.$$

(2) *If* $\operatorname{char}(K) = p \neq 0$*, then either*

$$\begin{aligned} E[p^e] &= \{O\} && \textit{for } e = 1, 2, 3, \ldots, \quad \textit{or} \\ E[p^e] &\cong \mathbb{Z}/p^e\mathbb{Z} && \textit{for } e = 1, 2, 3, \ldots. \end{aligned}$$

In the first case E is said to be supersingular, in the second case E is ordinary.

DEF supersingular ordinary

Proof.

(1) In this case $[m]$ is separable, which implies $\#E[m] = \deg([m]) = m^2$. In the same way, we have $\#E[d] = d^2$ for all divisors d of m. This, together with the structure theorem for finite abelian groups, implies the claim.

(2) Let $\phi\colon E \to E^{(p)}$, $(x, y) \mapsto (x^p, y^p)$ and let $\hat{\phi}\colon E^{(p)} \to E$ be the dual isogeny. Then we have (using that $\deg_s \phi = 1$; see Lemma 11.16, (1))

$$\#E[p^e] = \deg_s[p^e] = (\deg_s[p])^e = (\deg_s \hat{\phi}\phi)^e = (\deg_s \hat{\phi})^e\,.$$

Furthermore, $\deg_s \hat{\phi}$ is a divisor of $\deg \hat{\phi} = \deg \phi = p$. The two cases in the statement correspond to the two possibilities $\deg_s \hat{\phi} = 1$ and $\deg_s \hat{\phi} = p$. ❑

If E is an elliptic curve over a finite field K, then $E(K)$ is finite, say with $\#E(K) = n$, hence contained in $E[n]$. The structure theorem on finite abelian groups and the result above then imply that $E(K) \cong \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$, where $d_1 \mid d_2$ and $d_1 d_2 = n$. In addition, we must have $p \nmid d_1$ for $p = \operatorname{char}(K)$. In the following, we will describe an additional structure on $E[n]$ that reduces the possibilities for d_1 even further.

12.2. Theorem. *Let E be an elliptic curve over K. Then for every natural number m that is not a multiple of* $\operatorname{char}(K)$*, there exists a map*

THM Weil pairing

$$e_m\colon E[m] \times E[m] \to \mu_m$$

(where μ_m denotes the group of mth roots of unity in $\bar{K}$) with the following properties.

(1) *e_m is bilinear:*

$$e_m(S_1 + S_2, T) = e_m(S_1, T)e_m(S_2, T)\,,\ e_m(S, T_1 + T_2) = e_m(S, T_1)e_m(S, T_2)\,.$$

(2) *e_m is alternating: $e_m(T, T) = 1$. This implies that e_m is anti-symmetric: $e_m(T, S) = e_m(S, T)^{-1}$.*

(3) *e_m is non-degenerate: If $e_m(S, T) = 1$ for all $S \in E[m]$, then $T = O$. In particular, e_m is surjective.*

(4) *e_m is compatible with the action of the automorphism group of $\bar{K}$ over K, i.e., for $\sigma \in \operatorname{Aut}(\bar{K}/K)$ we have*

$$e_m(\sigma(S), \sigma(T)) = \sigma(e_m(S, T)) .$$

(5) *e_m and $e_{mm'}$ are compatible: For $S \in E[mm']$ and $T \in E[m]$,*

$$e_{mm'}(S, T) = e_m(m'S, T) .$$

(6) *If $\phi \colon E \to E'$ is an isogeny, then ϕ and $\hat{\phi}$ are adjoint with respect to e_m, i.e, for $S \in E[m]$ and $T \in E'[m]$,*

$$e_m(S, \hat{\phi}(T)) = e_m(\phi(S), T)$$

(where the e_m on the left belongs to E and the one on the right to E').

This map e_m is the *(m-)Weil pairing*.

DEF Weil pairing

A. Weil
1906–1998
Foto © MFO

Proof. Compare [Si1, § III.8]. Let $T \in E[m]$. By the Abel-Jacobi Theorem 10.15 there exists a rational function f_T on E such that $\operatorname{div}(f_T) = m \cdot (T) - m \cdot (O)$. Let $Q \in E$ be such that $mQ = T$. Then (in the group $E(\bar{K})$)

$$\sum_{R \in E[m]} (Q + R) = m^2 Q + \sum_{R \in E[m]} R = \sum_{R \in E[m]} R$$

(note that $m^2 Q = mT = O$), so again by Theorem 10.15 there is a rational function g_T on E such that

$$\operatorname{div}(g_T) = \sum_{R \in E[m]} \bigl((Q + R) - (R)\bigr) .$$

We then have $\operatorname{div}(g_T^m) = \operatorname{div}(f_T \circ [m])$; this implies that the quotient of these two functions is constant. By scaling f_T appropriately, we can assume that $f_T \circ [m] = g_T^m$.

Let now $S \in E[m]$. We consider the function

$$E \ni P \longmapsto \frac{g_T(P + S)}{g_T(P)} .$$

If $g_T(P)$ and $g_T(P + S)$ are both defined and $\neq 0$, then

$$\Bigl(\frac{g_T(P + S)}{g_T(P)}\Bigr)^m = \frac{g_T(P + S)^m}{g_T(P)^m} = \frac{f_T(mP + mS)}{f_T(mP)} = \frac{f_T(mP)}{f_T(mP)} = 1$$

(since $mS = O$). Hence the function above is constant, and its value is an mth root of unity. We define

$$e_m(S, T) = \frac{g_T(P + S)}{g_T(P)}$$

for every point $P \in E$ such that the right hand side is defined.

(1) For $S_1, S_2 \in E[m]$ we have with a suitable $P \in E$

$$\begin{aligned} e_m(S_1, T) e_m(S_2, T) &= \frac{g_T(P + S_1)}{g_T(P)} \, \frac{g_T((P + S_1) + S_2)}{g_T(P + S_1)} \\ &= \frac{g_T(P + (S_1 + S_2))}{g_T(P)} = e_m(S_1 + S_2, T) . \end{aligned}$$

For the other relation let h be a function with divisor

$$\operatorname{div}(h) = (T_1) + (T_2) - (O) - (T_1 + T_2) .$$

Then $f_{T_1+T_2}h^m = cf_{T_1}f_{T_2}$ with a constant $c \neq 0$. This implies $g_{T_1+T_2} \cdot (h \circ [m]) = c' g_{T_1} g_{T_2}$ and then

$$\begin{aligned} e_m(S,T_1)e_m(S,T_2) &= \frac{g_{T_1}(P+S)}{g_{T_1}(P)} \frac{g_{T_2}(P+S)}{g_{T_2}(P)} \\ &= \frac{g_{T_1+T_2}(P+S)h(mP+mS)}{g_{T_1+T_2}(P)h(mP)} = e_m(S, T_1+T_2)\,, \end{aligned}$$

since $mS = O$. (This also follows from the first relation and (2).)

(2) Let $Q \in E(\bar{K})$ such that $mQ = T$. The product

$$f_T(P)f_T(P+T)f_T(P+2T)\cdots f_T(P+(m-1)T)$$

is constant (since all zeros and poles cancel). This implies that the function

$$P \mapsto g_T(P)g_T(P+Q)g_T(P+2Q)\cdots g_T(P+(m-1)Q)$$

is also constant (since its mth power is essentially the previous product). Considering it for $P+Q$ instead of P, we obtain

$$\begin{aligned} &g_T(P)g_T(P+Q)g_T(P+2Q)\cdots g_T(P+(m-1)Q) \\ &\qquad = g_T(P+Q)g_T(P+2Q)\cdots g_T(P+(m-1)Q)g_T(P+mQ)\,, \end{aligned}$$

and so $g_T(P) = g_T(P+mQ) = g_T(P+T)$, which gives $e_m(T,T) = 1$.

(3) Let $T \in E[m]$ such that $e_m(S,T) = 1$ for all $S \in E[m]$. Then $g_T \circ \tau_S = g_T$ for all $S \in E[m]$. By Lemma 11.18, this implies that $g_T = h \circ [m]$ with some rational function h on E. Then

$$f_T \circ [m] = g_T^m = (h \circ [m])^m = h^m \circ [m]\,,$$

so $f_T = ch^m$ with a constant $c \neq 0$. Then we must have $\operatorname{div}(h) = (T) - (O)$, which is only possible when $T = O$.

The surjectivity of e_m then follows. The image of e_m is a subgroup of μ_m, so it is of the form μ_d for some divisor d of m. Then for all $S, T \in E[m]$ we have that $e_m(dS,T) = e_m(S,T)^d = 1$. The non-degeneracy of e_m then implies that $dS = O$ for all $S \in E[m]$, hence that $E[m] \subset E[d]$. But this implies that $d = m$.

(4) When we have fixed g_T, we can set $g_{\sigma(T)} = \sigma(g_T)$. Then we have

$$e_m(\sigma(S), \sigma(T)) = \frac{g_{\sigma(T)}(\sigma(P)+\sigma(S))}{g_{\sigma(T)}(\sigma(P))} = \sigma\Big(\frac{g_T(P+S)}{g_T(P)}\Big) = \sigma(e_m(S,T))\,.$$

(5) This is not hard (exercise).

(6) Let $Q \in E$ such that $\phi(Q) = T$. Then we have in $E(\bar{K})$

$$\sum_{R \in E[\phi]} ((Q+R) - R) = (\deg\phi)Q = \hat{\phi}(\phi(Q)) = \hat{\phi}(T)\,,$$

so by the Abel-Jacobi Theorem 10.15 there is a rational function h on E such that

$$\operatorname{div}(h) = \sum_{R \in E[\phi]} \big((Q+R) - (R)\big) - (\hat{\phi}(T)) + (O)\,.$$

Then (with $f_T, g_T \in \bar{K}(E')$ as above)

$$\operatorname{div}\Big(\frac{f_T \circ \phi}{h^m}\Big) = m \sum_{R \in E[\phi]} \big((Q+R) - (R)\big) - m\operatorname{div}(h) = m \cdot (\hat{\phi}(T)) - m \cdot (O)$$

and

$$\Bigl(\frac{g_T \circ \phi}{h \circ [m]_E}\Bigr)^m = \frac{f_T \circ [m]_{E'} \circ \phi}{(h \circ [m]_E)^m} = \frac{f_T \circ \phi}{h^m} \circ [m]_E \,.$$

We can therefore take $g_{\hat{\phi}(T)} = (f_T \circ \phi)/(h \circ [m])$. This gives

$$\begin{aligned} e_m(S, \hat{\phi}(T)) &= \frac{g_{\hat{\phi}(T)}(P+S)}{g_{\hat{\phi}(T)}(P)} = \frac{((g_T \circ \phi)/(h \circ [m]))(P+S)}{((g_T \circ \phi)/(h \circ [m]))(P)} \\ &= \frac{g_T(\phi(P) + \phi(S))}{g_T(\phi(P))} \cdot \frac{h(mP)}{h(mP+mS)} = e_m(\phi(S), T) \,. \end{aligned}$$ ❑

COR Structure of torsion

12.3. **Corollary.** *Let E be an elliptic curve over K. Let $\mu(K)$ be the subgroup of $K^\times$ consisting of all roots of unity in K. We assume that $\mu(K)$ is finite.*

Let G be a finite subgroup of $E(K)$. Then $G \cong \mathbb{Z}/d_1\mathbb{Z} \times \mathbb{Z}/d_2\mathbb{Z}$ with $d_1 \mid d_2$ and $d_1 d_2 = \#G$, where d_1 is a divisor of $\#\mu(K)$ and is not divisible by char(K).

Proof. Let $n = \#G$. Then $G \subset E[n]$ and $E[n] \subset \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, which already implies that G has the indicated shape; so we only have to prove the divisibility statements on d_1. If d_1 were a multiple of char(K), then char$(K) = p$ is a prime number, and we would have

$$\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z} \subset G \cap E[p] \subset E[p] \,,$$

which contradicts Theorem 12.1. For the claim that $d_1 \mid \#\mu(K)$, we note that $E[d_1] \subset G \subset E(K)$. Since the Weil pairing e_{d_1} is surjective, there are $S, T \in E(K)[d_1] = E[d_1]$ such that $e_{d_1}(S,T) = \zeta$ with $\zeta \in \bar{K}$ a primitive d_1th root of unity. Let $L = K(\zeta)$; this is a Galois extension of K. If we apply an element σ of the Galois group Gal(L/K) to the equation $e_{d_1}(S,T) = \zeta$, then the left hand side is unchanged, since S and T are defined over K. Since e_{d_1} is compatible with the Galois action, this implies that ζ also remains fixed. So $\zeta \in \mu(K)$ and therefore $d_1 = \operatorname{ord}(\zeta) \mid \#\mu(K)$. ❑

The statement of the corollary is analogous to the well-known fact that a finite subgroup of the multiplicative group of a field is always cyclic.

When E is an elliptic curve over $\mathbb{Q}$, the group $E(\mathbb{Q})$ is finitely generated (Mordell's Theorem; see later). This groups then has the form $E(\mathbb{Q}) \cong T \oplus \mathbb{Z}^r$, where T is some finite abelian group (and $r \in \mathbb{Z}_{\ge 0}$). Since $\mu(\mathbb{Q}) = \{\pm 1\}$, we find that T is either cyclic or has the form $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2d\mathbb{Z}$. A famous result due to Mazur then shows that when T is cyclic, the possibilities are exactly $\#T \le 10$ or $= 12$. In the other case, we must have $d \le 4$.

13. Elliptic curves over finite fields

We have already seen some specifics of elliptic curves over finite fields (or, at least, fields of nonzero characteristic). We will now consider elliptic curves over finite fields in more detail. This is motivated by several interesting applications.

We start by reminding ourselves of the most important facts about finite fields.

13.1. Theorem.

THM
finite fields

(1) *The number of elements of a finite field is a prime power p^f (with $f \geq 1$).*

(2) *Conversely, for every prime power $q = p^f$, there exists up to isomorphism exactly one finite field $\mathbb{F}_q$ with q elements.*

(3) *All extensions of finite fields have the form $\mathbb{F}_q \subset \mathbb{F}_{q^n}$; such a field extension is Galois with cyclic Galois group of order n. This group is generated by the Frobenius automorphism $x \mapsto x^q$.*

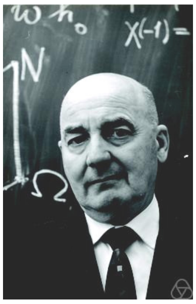
DEF
Frobenius automorphism

(4) *The algebraic closure of $\mathbb{F}_q$ is obtained as the filtered union $\bar{\mathbb{F}}_q = \bigcup_n \mathbb{F}_{q^n}$. Then*
$$\mathbb{F}_q = \{x \in \bar{\mathbb{F}}_q \mid x^q = x\}.$$

(5) *We have*
$$\mathbb{F}_q^\times = \{x \in \bar{\mathbb{F}}_q \mid x^{q-1} = 1\} = \mu_{q-1}(\mathbb{F}_q).$$
(where $\mu_n(K)$ denotes the group of nth roots of unity in K.)

Elliptic curves over finite fields have (at least) two prominent properties. On the one hand, the group of rational points is necessarily finite, and so its order is an important quantity. On the other hand, such a curve always has the Frobenius endomorphism in addition to the multiplication maps. We will soon see that these two things are connected.

A heuristic consideration leads one to expect that an elliptic curve over the finite field $\mathbb{F}_q$ should have about $q + 1$ rational points. The average number of solutions of an equation $y^2 = a$, where a runs through $\mathbb{F}_q$, is 1 (we assume that the characteristic is odd). If we assume that the values of a polynomial $f(x) = x^3 + a_2x^2 + a_4x + a_6$ are essentially randomly distributed, then the number of solutions of $y^2 = f(x)$ should be roughly q. Adding the point at infinity, we expect therefore $\#E(\mathbb{F}_q) \approx q + 1$, if E is the elliptic curve defined by the equation.



H. Hasse
1898 – 1979
Foto © MFO

This is indeed correct, and we can even bound the error quite precisely. The following theorem was first proved by Hasse.

13.2. Theorem. *Let E be an elliptic curve over the finite field $\mathbb{F}_q$ and let $\phi \in \mathrm{End}_{\mathbb{F}_q}(E)$ be its Frobenius endomorphism $(x, y) \mapsto (x^q, y^q)$.*

THM
Frobenius and number of points

DEF
trace of Frobenius

(1) *Let $t = \phi + \hat{\phi} \in \mathbb{Z}$ be the trace of Frobenius. Then we have in $\mathrm{End}_{\mathbb{F}_q}(E)$ the relation*
$$\phi^2 - t\,\phi + q = 0,$$
and $|t| \leq 2\sqrt{q}$.

(2) *We have $\#E(\mathbb{F}_q) = \deg(\phi - 1) = q + 1 - t$. In particular,*
$$|\#E(\mathbb{F}_q) - (q + 1)| \leq 2\sqrt{q}.$$

Proof.

(1) We compute in $\mathrm{End}_{\mathbb{F}_q}(E)$:

$$0 = (\phi - \phi)(\phi - \hat{\phi}) = \phi^2 - (\phi + \hat{\phi})\phi + \phi\hat{\phi} = \phi^2 - t\,\phi + q\,,$$

since $\phi\hat{\phi} = \deg(\phi) = q$.

For a rational number $r/s \in \mathbb{Q}$,

$$\Bigl(\frac{r}{s}\Bigr)^2 - t\,\frac{r}{s} + q = \frac{1}{s^2}(r^2 - t\,rs + q\,s^2) = \frac{1}{s^2}\deg(r - s\phi) \ge 0\,,$$

so the polynomial $X^2 - t\,X + q$ must have non-positive discriminant: $t^2 - 4q \le 0$, i.e., $|t| \le 2\sqrt{q}$.

(2) We have

$$\begin{aligned} E(\mathbb{F}_q) &= \{(\xi,\eta) \in E(\bar{\mathbb{F}}_q) \mid \xi = \xi^q, \eta = \eta^q\} \cup \{O\} \\ &= \{P \in E(\bar{\mathbb{F}}_q) \mid \phi(P) = P\} \\ &= \ker(\phi - 1)\,. \end{aligned}$$

Since $\phi - 1$ is separable (Lemma 11.16), this implies $\#E(\mathbb{F}_q) = \#\ker(\phi - 1) = \deg(\phi - 1)$ (Theorem 11.15). On the other hand,

$$\deg(\phi - 1) = (\phi - 1)(\hat{\phi} - 1) = \phi\hat{\phi} - (\phi + \hat{\phi}) + 1 = q - t + 1\,. \qquad ❑$$

R. Schoof
* 1955
Foto © MFO

N.D. Elkies
* 1966
Foto © MFO

One can ask whether there is a reasonably fast way of actually computing $\#E(\mathbb{F}_q)$ from a given equation of E. There is indeed an efficient (i.e., polynomial-time in $\log q$) algorithm that determines the number of rational points on an elliptic curve over $\mathbb{F}_q$. It was developed theoretically by Schoof and then made practical by Atkin and Elkies. The basic idea is to determine the residue class of $t \bmod \ell$ for sufficiently many suitable prime numbers ℓ and then obtain the value of t (and hence of $\#E(\mathbb{F}_q) = q + 1 - t$) from this via the Chinese Remainder Theorem (and the bound $|t| \le 2\sqrt{q}$).

Theorem 13.2 together with Corollary 12.3 leads to the following statement about the structure of the group $E(\mathbb{F}_q)$.

$$E(\mathbb{F}_q) \cong \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/dd'\mathbb{Z}\,,$$

where $d \mid q - 1$ and $|d^2d' - (q + 1)| \le 2\sqrt{q}$.

We now state a result on isogenies and the number of rational points.

13.3. Theorem. *Let E and E' be two elliptic curves over $\mathbb{F}_q$. Then the following two statements are equivalent.*

THM
isogenous
ell. curves

(1) *E and E' are isogenous over $\mathbb{F}_q$.*

(2) *$\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$.*

Proof. We will prove only the direction ‘(1) ⇒ (2)’ here. The proof of the converse requires quite deep results.

We suppose that we have a (non-constant) isogeny $\psi\colon E \to E'$ that is defined over $\mathbb{F}_q$. We denote the Frobenius endomorphisms of E and E' by ϕ and ϕ', respectively, and by t and t' their traces. Since the map $x \mapsto x^q$ commutes with the basic arithmetic operations and fixes the elements of $\mathbb{F}_q$, we have $\psi \circ \phi = \phi' \circ \psi$. In the same way, we get $\phi \circ \hat{\psi} = \hat{\psi} \circ \phi'$, which implies by dualizing that $\psi \circ \hat{\phi} = \hat{\phi}' \circ \psi$. Using both relations, we see that

$$\psi \circ [t] = \psi \circ \phi + \psi \circ \hat{\phi} = \phi' \circ \psi + \hat{\phi}' \circ \psi = [t'] \circ \psi = \psi \circ [t']\,.$$

(The last equation holds because ψ is a homomorphism.) Composing with $\hat{\psi}$ on the left then gives the equation

$$\deg(\psi)t = \deg(\psi)t'$$

in $\operatorname{End}(E)$. Since $\deg(\psi) \neq 0$ and $\operatorname{End}(E)$ is an integral domain of characteristic zero (Theorem 11.9), we see that $t = t'$ and therefore by Theorem 13.2 $\#E(\mathbb{F}_q) = \#E'(\mathbb{F}_q)$ as well. ❑

The zeta function. We have seen that the number of rational points on an elliptic curve E over $\mathbb{F}_q$ is closely related to the behavior of the Frobenius endomorphism ϕ. We can consider E also as an elliptic curve over $\mathbb{F}_{q^n}$ for $n = 2, 3, 4, \dots$. We will now study how the numbers

$$\#E(\mathbb{F}_q)\,, \quad \#E(\mathbb{F}_{q^2})\,, \quad \#E(\mathbb{F}_{q^3})\,, \quad \dots$$

are related. For this purpose, we introduce an object that encodes the information about all these numbers in a suitable form.

13.4. **Definition.** Let C be a smooth projective curve over $\mathbb{F}_q$. The *zeta function* of C is the following power series with rational coefficients.

DEF
Zeta function

$$\begin{aligned} Z(C,T) &= \exp\left(\#C(\mathbb{F}_q)\,T + \frac{\#C(\mathbb{F}_{q^2})}{2}\,T^2 + \frac{\#C(\mathbb{F}_{q^3})}{3}\,T^3 + \dots\right) \\ &= \exp\left(\sum_{n=1}^{\infty} \frac{\#C(\mathbb{F}_{q^n})}{n}\,T^n\right). \end{aligned} \qquad \diamondsuit$$

The connection with the variant $\sum_{n\ge1} \#C(\mathbb{F}_{q^n})T^n$ that one would perhaps first think of is given by the logarithmic derivative:

$$\sum_{n\ge1} \#C(\mathbb{F}_{q^n})T^n = T\,\frac{\frac{d}{dT}Z(C,T)}{Z(C,T)} = T\,\frac{d}{dT}\log Z(C,T)\,.$$

The reason for the at first sight somewhat contrived definition of the zeta function is that in this form it has a natural product expansion. To see this, we consider the set of algebraic points on C, $C(\bar{\mathbb{F}}_q) = \bigcup_{n\ge1} C(\mathbb{F}_{q^n})$. This set decomposes into orbits under the action of the Frobenius endomorphism ϕ. Let a_d be the number of orbits of length d. Then $\#C(\mathbb{F}_{q^n}) = \sum_{d|n} da_d$, and the zeta function can be written

$$Z(C,T) = \prod_{d=1}^{\infty}(1 - T^d)^{-a_d}$$

(Exercise). It also turns out that the zeta function as it is defined has a particularly simple form, as shown by the following result.

13.5. **Theorem.** *Let C be a smooth projective curve over $\mathbb{F}_q$.*

THM
Weil
conjectures
for curves

(1) *$Z(C,T) \in \mathbb{Q}(T)$ (i.e., $Z(C,T)$ is the power series of a rational function).*

(2) *$Z(C, 1/(qT)) = q^{1-g}T^{2-2g}Z(C,T)$ (Functional equation). Here g is the genus of C ($g = 1$ for elliptic curves).*

(3) *$Z(C,T) = P(T)/((1-T)(1-qT))$ with a polynomial $P(T) \in \mathbb{Z}[T]$ of degree $2g$ that splits over $\mathbb{C}$ as*

$$P(T) = \prod_{j=1}^{g} \big((1-\alpha_j T)(1-\bar{\alpha}_j T)\big)$$

with $|\alpha_j| = \sqrt{q}$. ("Riemann Hypothesis")

13.6. **Remarks.**

REM
Weil conjectures

(1) Weil has stated his conjectures more generally also for projective varieties of higher dimension. For curves (and abelian varieties) he proved them himself (1949). The various parts of the general conjecture were dealt with by Deligne between 1960 and 1973.

(2) The *genus* g is an important invariant of the curve C; however, it is not easy to define. For a smooth *plane* projective curve of degree d, we have $g = \frac{1}{2}(d-1)(d-2)$; so for elliptic curves (which are smooth plane projective curves of degree 3) we have $g = 1$.

(3) The designation "Riemann Hypothesis" for part (3) of the theorem is motivated by the following analogy. If we put $\zeta(C,s) = Z(C,q^{-s})$, then this function ζ has simple poles at $s = 0$ and $s = 1$, and all its zeros have real part $\frac{1}{2}$. (Furthermore, the functional equation translates to $\zeta(C,1-s) = q^{(g-1)(2s-1)}\zeta(C,s)$, which is similar to the functional equation of the Riemann zeta function.) ♠

We will now prove the theorem for elliptic curves.

Proof. Let E be an elliptic curve over $\mathbb{F}_q$ and let $\phi \in \operatorname{End}(E)$ be the Frobenius endomorphism. We had already seen that ϕ solves the equation $X^2 - t\,X + q = 0$ (Theorem 13.2), where $t = \phi + \hat{\phi}$ is the trace of Frobenius. We also have $|t| \le 2\sqrt{q}$, which implies that

$$X^2 - t\,X + q = (X-\alpha)(X-\bar{\alpha})$$

with $\alpha \in \mathbb{C}$, $|\alpha| = \sqrt{q}$. Since we similarly have

$$X^2 - t\,X + q = (X-\phi)(X-\hat{\phi})\,,$$

we obtain an isomorphism

$$\mathbb{Z}[\alpha] \longrightarrow \mathbb{Z}[\phi] \subset \operatorname{End}(E)\,, \qquad \alpha \longmapsto \phi\,.$$

(When $\alpha = \pm\sqrt{q}$, we use that $\operatorname{End}(E)$ is an integral domain; see Theorem 11.9.) Now,

$$\#E(\mathbb{F}_q) = q + 1 - \phi - \hat{\phi} = q + 1 - \alpha - \bar{\alpha}\,,$$

and then in a similar way (note that ϕ^n is the Frobenius endomorphism of E over $\mathbb{F}_{q^n}$)

$$\#E(\mathbb{F}_{q^n}) = q^n + 1 - \phi^n - \hat{\phi}^n = q^n + 1 - \alpha^n - \bar{\alpha}^n\,.$$

We deduce that

$$\begin{aligned}
Z(E,T) &= \exp\left(\sum_{n=1}^{\infty} \frac{\#E(\mathbb{F}_{q^n})}{n} T^n\right) \\
&= \exp\left(\sum_{n=1}^{\infty} \frac{(qT)^n}{n} + \sum_{n=1}^{\infty} \frac{T^n}{n} - \sum_{n=1}^{\infty} \frac{(\alpha T)^n}{n} - \sum_{n=1}^{\infty} \frac{(\bar{\alpha} T)^n}{n}\right) \\
&= \exp\left(\log \frac{1}{1-qT} + \log \frac{1}{1-T} - \log \frac{1}{1-\alpha T} - \log \frac{1}{1-\bar{\alpha} T}\right) \\
&= \frac{(1-\alpha T)(1-\bar{\alpha} T)}{(1-T)(1-qT)} = \frac{1 - t\,T + q\,T^2}{(1-T)(1-qT)}\,.
\end{aligned}$$

This shows part (1). Part (2) follows from a simple calculation, and part (3) is a consequence of $|\alpha| = \sqrt{q}$. ❑

Perhaps the most surprising consequence of this result is that the number of rational points over $\mathbb{F}_q$ on an elliptic curve E already determines the numbers $\#E(\mathbb{F}_{q^n})$ for *all* $n \geq 1$!

14. Factorization and primality proving: basics

After learning about elliptic curves in general and also about some special properties of elliptic curves over finite fields, we can now look at some practical applications. The first of these will be the factorization of large numbers and in connection with that the proof that a large number is prime. The main source for this and the following sections is [Co1]. But first we need to consider the problem in more detail.

A preliminary remark on factorization in practice: it is a recursive procedure that can be split into the following parts.

- Determine if a positive natural number N is composite or most likely prime ("Compositeness test").
- If N is likely to be prime, prove that N is indeed prime ("Primality proof").
- If N is composite, find a nontrivial factor d of N and continue with d and N/d ("Factorization" proper).

Usually one will first perform a trial division to find all sufficiently small prime divisors of N.

Compositeness test.

In order to show that a number is composite, one can test whether it satisfies conditions that are valid for all primes. One possibility here is *Fermat's Little Theorem*, which says that

$$a^{p-1} \equiv 1 \bmod p\,.$$

for all primes p and all integers a with $p \nmid a$. This leads to the following definition. We write "$a \perp N$" for the statement that a and N are coprime.

14.1. Definition. An integer $N > 1$ is a *pseudoprime to base a*, if $a^{N-1} \equiv 1 \bmod N$ (this implies $a \perp N$).

DEF pseudoprime

N is a *Carmichael number*, if N is not prime, but N is a pseudoprime to base a for all $a \perp N$. ♢

Carmichael number

A prime number is clearly a pseudoprime to all bases a with $p \nmid a$. So if we can find $1 < a < N$ such that $a^{N-1} \not\equiv 1 \bmod N$, then this shows that N is composite. In this context it is important that we can compute the residue of a^{N-1} modulo N efficiently (by successive squaring and repeated reduction mod N; using standard multiplication and division algorithms, this gives a complexity of $O((\log N)^3)$; fast multiplication and division leads essentially to $O((\log N)^2)$). Unfortunately, this test does not always work.

14.2. Theorem. (Alford, Granville, Pomerance 1994[1])
There are infinitely many Carmichael numbers.

THM infinitely many Carmichael numbers

If N is a Carmichael number, then we always have $a^{N-1} \equiv 1 \bmod N$, unless $\gcd(a, N) > 1$, which is extremely unlikely for a random choice of a when N is large.

There is, however, a variant that works better. This is achieved by sharpening the condition in Definition 14.1.

14.3. Definition. Let N be an odd natural number. We write $N - 1 = 2^t q$ with

DEF strong pseudoprime

[1] W. R. Alford, A. Granville, and C. Pomerance: *There are infinitely many Carmichael numbers,* Annals of Mathematics **139** (1994) 703–722.

q odd. Let further a be an integer. Then N is a *strong pseudoprime to base* a, if

$$a^q \equiv 1 \bmod N \qquad \text{or} \qquad a^{2^e q} \equiv -1 \bmod N \quad \text{for some } 0 \le e < t. \qquad \diamondsuit$$

The assumption that N is odd is no essential restriction, since we can very easily check whether N is divisible by 2.

14.4. **Theorem.** *Let N be an odd natural number.*

THM Miller-Rabin test

(1) *If N is prime, then N is a strong pseudoprime to base a for all a such that $N \nmid a$.*

(2) *If N is composite, then N is a strong pseudoprime to base a for fewer than $N/4$ numbers a such that $1 < a < N$.*

Proof.

(1) If $N = p$ is a prime, then $x^2 \equiv 1 \bmod p$ implies that $x \equiv \pm 1 \bmod p$ (since the polynomial $X^2 - 1$ can have at most two roots in the field $\mathbb{F}_p$). Fermat's Little Theorem says that $a^{p-1} = a^{2^t q} \equiv 1 \bmod p$. We can then conclude that we have either $a^q \equiv 1 \bmod p$ or $a^{2^e q} \equiv -1 \bmod p$ for some $0 \le e < t$.

(2) We first consider the homomorphism

$$(\mathbb{Z}/N\mathbb{Z})^\times \longrightarrow (\mathbb{Z}/N\mathbb{Z})^\times, \qquad \bar{a} \longmapsto \bar{a}^{N-1}.$$

Let $G \subset (\mathbb{Z}/N\mathbb{Z})^\times$ be its kernel. Then $\#G \le \#(\mathbb{Z}/N\mathbb{Z})^\times < N$. ($N$ is a Carmichael number iff $G = (\mathbb{Z}/N\mathbb{Z})^\times$.) Let further $N = p_1^{e_1} \cdots p_k^{e_k}$ be the prime factorization of N. Then the primes p_j are odd, and

$$(\mathbb{Z}/N\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{e_1}\mathbb{Z})^\times \times \cdots \times (\mathbb{Z}/p_k^{e_k}\mathbb{Z})^\times$$

is a product of cyclic groups of even order $(p_j - 1)p_j^{e_j - 1}$. This induces a corresponding splitting of G as $G \cong G_1 \times \cdots \times G_k$, where G_j is cyclic and has (even) order $\gcd(N - 1, (p_j - 1)p_j^{e_j - 1})$. Let $G'_j \subset G_j$ be the unique subgroup of index 2. If $a \in \mathbb{Z}$ is such that $\bar{a} \in G_j$, we then have: if $a \bmod p_j^{e_j}$ is in G'_j, then $a^{(N-1)/2} \equiv 1 \bmod p_j^{e_j}$, and otherwise, $a^{(N-1)/2} \equiv -1 \bmod p_j^{e_j}$. For $\bar{a} \in G$ we set $\varepsilon_j(\bar{a}) = 1$, if the image of $\bar{a}$ in G_j is in G'_j, otherwise $\varepsilon_j(\bar{a}) = -1$. Then

$$\varepsilon \colon G \longrightarrow \{\pm 1\}^k, \qquad \bar{a} \longmapsto (\varepsilon_1(\bar{a}), \ldots, \varepsilon_k(\bar{a}))$$

is a surjective group homomorphism. We have

$$a^{(N-1)/2} \equiv \pm 1 \bmod N \iff \varepsilon(\bar{a}) = \pm(1, \ldots, 1).$$

This implies

$$\#\{\bar{a} \in G \mid a^{(N-1)/2} \equiv \pm 1 \bmod N\} = 2^{1-k}\#G.$$

If N is neither a Carmichael number nor a prime power, then $\#G < N/2$ and $k \ge 2$, hence fewer than $2^{-k}N \le N/4$ numbers a satisfy the necessary condition

$$a^{(N-1)/2} \equiv \pm 1 \bmod N.$$

If N is a Carmichael number, then $k \ge 3$ (exercise), which again implies the claim. If finally $N = p^e$ is a prime power (with $e \ge 2$), then

$$\#G = \gcd(p^e - 1, (p-1)p^{e-1}) = p - 1 < p^e/4,$$

and the claim also follows. ❑

This result leads to the *Miller-Rabin test*.

14.5. **Algorithm.** (Miller-Rabin test)

```
Input: $N > 1$ odd; $m \geq 1$ (number of tests)
  Write $N - 1 = 2^t q$ with $q$ odd
  For $j = 1, \ldots, m$:
    Pick $1 < a < N$ randomly and compute $b := a^q \bmod N$
    If $b = \pm 1$: go to the next $j$
    For $e = 1, \ldots, t - 1$:
      Set $b := b^2 \bmod N$.
      If $b = -1$: go to the next $j$
    // (N is not a strong pseudoprime to base a)
    Print ‘$N$ is composite’ and stop
  // (N has passed all tests)
  Print ‘$N$ is probably prime’ and stop
```

The result of Theorem 14.4 (2) tells us that the probability that a composite number N is declared as ‘probably prime’ is less than 4^{-m}.

Primality proof.

On the other hand, it is impossible to *prove* that N is indeed prime in this way. One possibility to do this is to use a suitable converse of Fermat's Little Theorem.

14.6. **Lemma.** *Let $N > 0$ be an integer and let p be a prime divisor of $N - 1$. Let further $a_p \in \mathbb{Z}$ such that*

LEMMA
converse
of FIT

$$a_p^{N-1} \equiv 1 \bmod N \qquad \text{and} \qquad (a_p^{(N-1)/p} - 1) \perp N\,. \tag{14.1}$$

Let p^{e_p} be the highest power of p dividing $N - 1$. Then every (positive) divisor d of N satisfies

$$d \equiv 1 \bmod p^{e_p}\,.$$

Proof. We can restrict consideration to prime divisors d. Since $a_p \perp N$, we also have $d \nmid a_p$ and hence $a_p^{d-1} \equiv 1 \bmod d$. On the other hand, $a_p^{(N-1)/p} \not\equiv 1 \bmod d$, since $a_p^{(N-1)/p} - 1$ and N are coprime by assumption. Denote the order of a_p mod d by n; then we have $n \mid d - 1$, $n \mid N - 1$ (since $a_p^{N-1} \equiv 1 \bmod d$), but $n \nmid (N-1)/p$. The last two properties imply that $p^{e_p} \mid n$, and the first then gives that $p^{e_p} \mid d - 1$. ❑

If we know the factorization of $N - 1$ sufficiently well, then we can use this result to show that N is prime.

14.7. **Corollary.** *Let $N > 0$ be an integer such that $N - 1 = F \cdot U$ with $F \geq \sqrt{N}$; we assume that all prime divisors of F are known.*

COR
Pocklington-
Lehmer test

N is prime, if and only if for each prime divisor p of F there is a number $a_p \in \mathbb{Z}$ that satisfies (14.1).

Proof. First assume that N is prime. Let g be a primitive root mod N (i.e., the image of g generates that group $(\mathbb{Z}/N\mathbb{Z})^\times$). Then $a_p = g$ satisfies (14.1).

Now assume that for each $p \mid F$ we have some a_p satisfying (14.1). Lemma 14.6 then implies that every divisor d of N satisfies the congruence $d \equiv 1 \bmod F$. In particular, $d = 1$ or $d > F \geq \sqrt{N}$. If N were composite, then N would have a nontrivial divisor $\leq \sqrt{N}$, but we have just excluded this possibility. Therefore N must be prime. ❑

One can obtain a method for proving primality from this, the *Pocklington-Lehmer test*. It is based on the use of the cyclic group $(\mathbb{Z}/N\mathbb{Z})^\times$ of order $N-1$. Its disadvantage is that it requires a good (partial) knowledge of the factorization of $N-1$, which can be a considerable obstacle in practice. As an aside, this also shows that it is frequently necessary to factor numbers, if one wants to prove that a certain number is prime. This reinforces the recursive nature of factorization.

One can modify this approach by using the subgroup of order $N+1$ of $\mathbb{F}_{N^2}^\times$ in place of $\mathbb{F}_N^\times$. Then one needs information on the factorization of $N+1$. This leads, for example, to the well-known *Lucas-Lehmer test* for Mersenne primes $2^p - 1$.

Elliptic curves can help here, since they provide groups of order roughly N, but such that the group orders have enough variation so that there is a good chance of finding a group with a sufficiently factorizable order. We will discuss this in more detail in the next section.

A discussion of algorithms for primality proving would be incomplete without mentioning the deterministic polynomial-time algorithm of Agrawal, Kayal and Saxena.[2] This results solves an old problem, since up to then no method was known that would decide for an arbitrary natural number deterministically (i.e., without random choices like in the Miller-Rabin test) and in polynomial time whether it is prime or not. This breakthrough was obtained in the context of a bachelor's project of the two (then) students Kayal and Saxena under Agrawal's supervision. This is bases on a generalization of Fermat's Little Theorem to polynomials, which provides a characterization of prime numbers: for every integer $a \perp N$, we have the equivalence

$$N \text{ is prime} \iff (X-a)^N \equiv X^N - a \bmod N$$

in the polynomial ring $\mathbb{Z}[X]$ (i.e., the congruence mod N holds for each coefficient). The computation of the left hand side is much too involved, however. This is why one considers instead the congruence

$$(X-a)^N \equiv X^N - a \bmod \langle N, X^r - 1 \rangle$$

for suitable $r \geq 1$. The three authors were able to show that the validity of this congruence for r and a as in the algorithm below is sufficient to imply that N is a prime power.

14.8. **Algorithm.** (AKS primality test)

Input: $N > 1$.

```
If $N$ is a proper power, print ‘composite’; stop.
Find the smallest $r \geq 1$ such that $\operatorname{ord}_r(N) > (\log_2 N)^2$.
If $1 < \gcd(a, N) < N$ for some $1 \leq a \leq r$, print ‘composite’; stop.
If $N \leq r$, print ‘prime’; stop.
For $a = 1, \ldots, \lfloor \sqrt{\varphi(r)} \log_2 N \rfloor$:
  If $(X-a)^N \not\equiv X^N - a \bmod \langle N, X^r - 1 \rangle$:
    print ‘composite’; stop.
print ‘prime’; stop.
```

Here $\operatorname{ord}_r(N)$ denotes the order of N in the multiplicative group $(\mathbb{Z}/r\mathbb{Z})^\times$, and $\varphi(r)$ is the Euler φ function, i.e., the order of this group.

They could also show that r is sufficiently small to ensure that the running time can be bounded by a polynomial in $\log N$ (originally, the bound was $O((\log N)^{12})$, but it has been improved since).

However, probabilistic algorithms like the one we will describe in the next section are sill faster in practice.

[2] Manindra Agrawal, Neeraj Kayal, Nitin Saxena. *PRIMES is in P*, Annals of Mathematics **160** (2004), no. 2, 781–793.

Factorization.

After we have discussed how one can check whether a number is prime, we now consider factorization. We assume that we are given a number N and we know that it is composite (for example, from one of the compositeness tests). The goal is to find a nontrivial divisor d of N.

14.9. **Definition.** We say that a positive integer is *B-smooth*, if all its prime divisors are $\leq B$. The number is *B-powersmooth*, if all prime powers that divide it are $\leq B$. ♢

DEF
B-smooth
B-powersmooth

We have seen that the Pocklington-Lehmer test requires a kind of smoothness condition for $N-1$. The factorization algorithm we will now describe has a similar restriction: it can only find divisors when there are prime divisors p of N such that $p-1$ is B-powersmooth.

The idea is as follows. We pick a bound B and an integer a. If N has a prime divisor p such that $p-1$ is B-powersmooth, then $p-1$ divides $L(B) = \operatorname{lcm}(1, 2, \ldots, B)$. Fermat's Little Theorem implies that $a^{L(B)} \equiv 1 \bmod p$, hence

$$\gcd(a^{L(B)} - 1, N) > 1\,.$$

This gcd therefore is a divisor > 1 of N, and with some luck, it is $< N$ as well (and so a nontrivial divisor). In practice, one computes successively $a^{L(1)} \bmod N$, $a^{L(2)} \bmod N$, …, $a^{L(B)} \bmod N$ (by successively computing powers mod N with exponent $L(n+1)/L(n)$; this is either 1 or a prime number q) and checks the gcd each time $L(n)$ has changed.

This algorithm is due to Pollard (who found several other factorization algorithms as well). The choice of B mainly depends on how much effort we are prepared to invest.

14.10. **Example.** We consider $N = 119$. As a first step, we determine that N is composite: $N - 1 = 118 = 2 \cdot 59$, and using $a = 2$ in the Miller-Rabin test, we find that $a^{59} \equiv 25 \bmod 119$ and $a^{118} \equiv 30 \bmod 119$, so that N fails the test.

EX
Pollard
$p-1$

Now we want to find a divisor of N. We pick again $a = 2$. We obtain

$$a^{L(2)} = a^2 \equiv 4 \bmod 119, \qquad \gcd(3, 119) = 1\,,$$
$$a^{L(3)} = a^6 \equiv 64 \bmod 119, \qquad \gcd(63, 119) = 7\,,$$

and we have found a divisor: $119 = 7 \cdot 17$. ♣

This algorithm can also be modified to use a group of order $p+1$; then one can find divisors p such that $p+1$ is B-powersmooth. If one insists on working with the multiplicative group of a field, then one is essentially confined to these two possibilities, if one does not want to use significantly larger groups (of size about p^2 or larger).

At this point, elliptic curves again can help, since an elliptic curve over $\mathbb{F}_p$ also provides us with an abelian group of size roughly p, but with the order varying in an interval around $p+1$, so that there is a good chance of finding a B-powersmooth number in this range.

Before we discuss how elliptic curves can be used, I would like to mention some further factorization methods.

One of them is based on the *birthday paradox.* The idea is as follows. Let N be the number to be factored. We consider a function $f\colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/N\mathbb{Z}$ that is easy to evaluate, for example $f(x) = x^2 + 1$. We assume that f behaves with respect to iteration like a random map. We pick $x_0 \in \mathbb{Z}/N\mathbb{Z}$ and compute $x_1 = f(x_0)$, $x_2 = f(x_1)$, and so on. For p a prime divisor of N, the sequence $(x_n \bmod p)$ will be eventually periodic; so there will by n and $m \geq 1$ such that $x_{n+m} \equiv x_n \bmod p$. With some luck, this relation does not hold for all prime divisors of N, so that we obtain a nontrivial divisor of N from $\gcd(x_{n+m} - x_n, N)$.

To keep the number of comparisons reasonable, we can compute the sequence (x_{2n}) in parallel to (x_n) and then compute $\gcd(x_{2n} - x_n, N)$ for every n. (Further improvements are possible.) One can expect (but this is not guaranteed) to find a divisor in time $O(\sqrt{p}(\log N)^2)$, when p is the smallest prime divisor of N.

14.11. **Example.** We again consider $N = 119$. We take $f(x) = x^2 + 1$ and $x_0 = 1$. We compute

EX Pollard ρ

n	0	1	2	3	4
x_n	1	2	5	26	82
$x_n \bmod 7$	1	2	5	5	5
$x_n \bmod 17$	1	2	5	9	14

and find $\gcd(x_2 - x_1, 119) = 1$, $\gcd(x_4 - x_2, 119) = 7$. ♣

Most modern factorization methods (but this is not true of the method using elliptic curves) rely on the following idea. Let N be an odd composite natural number that has at least two distinct prime factors. (It is easy to check whether N is a power, so this is no restriction.) If we find two integers x and y such that $x^2 \equiv y^2 \bmod N$, then with probability $\geq 1/2$ (assuming that x and y were chosen randomly in the residue classes mod N), $\gcd(x - y, N)$ will be a nontrivial divisor of N. The reason is that for every prime divisor p of N we have $x \equiv \varepsilon_p y \bmod p^{v_p(N)}$ with $\varepsilon_p = \pm 1$. These signs are independent from one another, and we obtain a nontrivial divisor as soon as not all signs are equal (compare the proof of Theorem 14.4).

One therefore tries to generate congruences of the form $x^2 \equiv y^2 \bmod N$. In order to do this, one fixes a bound B and considers the prime numbers $q_1, q_2, \ldots, q_k \leq B$. The set $\{-1, q_1, \ldots, q_k\}$ is known as the *factor base*. One then tries to obtain relations of the form

$$x^2 \equiv (-1)^{e_0} q_1^{e_1} \cdots q_k^{e_k} \bmod N .$$

Once one has collected sufficiently many of those, one can (using linear algebra over $\mathbb{F}_2$) find subsets of these relations such that the product of the right hand sides is a square. The product of the left hand sides is a square in any case, and so one obtains a relation of the desired form. The various method differ in the way they use to generate the original relations.

One method uses continued fractions. For $k = 1, 2, \ldots$ one computes the beginning of the continued fraction expansion of $\sqrt{kN}$ and from this the first few approximating fractions r/s. Then $t = r^2 - s^2 kN$ is relatively small, so that one can hope that t factors over the factor base. Note that $r^2 \equiv t \bmod N$.

14.12. **Example.** Let again $N = 119$. The first approximating fractions for $\sqrt{119}$ are

EX factorization with continued fractions

$$\frac{10}{1}, \quad \frac{11}{1}, \quad \frac{109}{10}, \quad \ldots$$

We obtain the relations

$$\begin{aligned} 10^2 &\equiv -19 = (-1)\cdot 19 && \bmod 119 \\ 11^2 &\equiv 2 = 2 && \bmod 119 \end{aligned}$$

(the ones after that do not provide new information). For $\sqrt{2\cdot 119}$ we find the approximations 15 and 31/2, leading to

$$\begin{aligned} 15^2 &\equiv -13 = (-1)\cdot 13 && \bmod 119 \\ 31^2 &\equiv 9 = 3^2 && \bmod 119\,, \end{aligned}$$

and this final relation gives the divisor $\gcd(31-3, 119) = 7$. ♣

The *quadratic sieve* uses polynomials like

$$Q(x) = (\lfloor\sqrt{N}\rfloor + x)^2 - N$$

to produce relatively small number that are congruent mod N to squares. For the (in general fairly expensive) factorization of these numbers one can use that the divisibility of $Q(a)$ by p depends only on the residue class of a mod p. This gives a fast way of removing the primes in the factor basis very quickly from all values $Q(a)$ with $-B < a < B$, and one can determine those that factor completely.

This method, when optimized (“Multiple Polynomial Quadratic Sieve”, MPQS), has an expected running time of the order $O(e^{c\sqrt{\log N \log\log N}})$; it is currently one of the best available methods. The *number field sieve*, which uses similar ideas, but works in an algebraic number field, even has a (conjectured) complexity of $O(e^{c\sqrt[3]{\log N(\log\log N)^2}})$. Due to the more complicated computations, it becomes faster only in a range that is close to the boundary of the feasible.

15. Factorization and primality proving with elliptic curves

To be able to formulate the results below in a reasonable way, we need the notion of an elliptic curve over $\mathbb{Z}/N\mathbb{Z}$. We can more generally consider elliptic curves over a (commutative) ring R (with 1). They can be defined in the same way as over a field. To talk about their points, we need to figure out how to define the R-points of the projective plane. It turns out that the correct definition is as follows.

$$\mathbb{P}^2(R) = \{(\xi, \eta, \zeta) \in R^3 \mid R \cdot \xi + R \cdot \eta + R \cdot \zeta = R\}/\sim\,,$$

where the equivalence relation $\sim$ is given again by

$$(\xi, \eta, \zeta) \sim (\xi', \eta', \zeta') \iff \exists \lambda \in R^\times : (\xi', \eta', \zeta') = \lambda \cdot (\xi, \eta, \zeta)\,.$$

The main point here is that "$\neq 0$" is replaced by "invertible" or "coprime". Using this definition, most notion carry over. An elliptic curve over R is then given by a long Weierstrass equation with coefficients in R such that the discriminant is invertible.

We observe that a ring homomorphism $\phi\colon R \to S$ induces a map $\mathbb{P}^2(R) \to \mathbb{P}^2(S)$ that is compatible with all constructions. In particular, when we have an elliptic curve E over R, then applying ϕ to the coefficients of the equation defining E gives the equation of an elliptic curve E' over S, and we obtain a map $E(R) \to E'(S)$. (We have already used this implicitly in the case that $R \to S$ is a field extension.)

We will apply this to $R = \mathbb{Z}/N\mathbb{Z}$. Since we can always remove small prime divisors by trial division, we can assume $6 \perp N$, i.e., that 2 and 3 are invertible in $\mathbb{Z}/N\mathbb{Z}$. In this case, we can again transform a long Weierstrass equation into a short one

$$E\colon y^2 = x^3 + a\,x + b$$

with $a, b \in \mathbb{Z}/N\mathbb{Z}$ such that $4a^3 + 27b^2 \in (\mathbb{Z}/N\mathbb{Z})^\times$.

What is less obvious is whether $E(\mathbb{Z}/N\mathbb{Z})$ is still a group in a natural way. We will pretend that it is and that the formulas for the computation of sums and multiplies of points that work over a field are still valid. This uses only the four basic arithmetic operations, so the only thing that can go wrong is that we are supposed to divide by a nonzero element a that is not invertible. In this case the computation of the gcd of a and N that is necessary to find the inverse of a will yield a nontrivial divisor of N, and we are done (since we have found a divisor and therefore also have shown that N is not prime). Therefore we can assume that all computations can be done in the usual way.

Primality proof.

We begin by considering the problem to show that N is prime. The following result is analogous to Lemma 14.6.

15.1. **Lemma.** *Let $N > 1$ be an integer such that $6 \perp N$, and let E be an elliptic curve over $\mathbb{Z}/N\mathbb{Z}$. Let further $P \in E(\mathbb{Z}/N\mathbb{Z})$, $m \in \mathbb{Z}$ and $q > (\sqrt[4]{N} + 1)^2$ a prime divisor of m such that*

LEMMA primality criterion

$$m \cdot P = O \qquad \text{and} \qquad \frac{m}{q} \cdot P = (\xi : \eta : \zeta) \text{ with } \zeta \in (\mathbb{Z}/N\mathbb{Z})^\times. \tag{15.1}$$

Then N is prime.

Proof. Assume that N is not prime; then there is a prime $p \mid N$ such that $p \le \sqrt{N}$. The canonical homomorphism $\mathbb{Z}/N\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z} = \mathbb{F}_p$ turns E into an elliptic curve E' over $\mathbb{F}_p$; let P' be the image of P and let n be its order in $E'(\mathbb{F}_p)$. Then $n \mid m$ (since $mP' = O$), but $n \nmid m/q$ (since $(m/q)P' \ne O$ by assumption; note that the image of ζ in $\mathbb{F}_p$ is nonzero). This implies that $q \mid n$. We obtain the contradictory chain of inequalities

$$q \le n \le \#E'(\mathbb{F}_p) \le p + 1 + 2\sqrt{p} = (1+\sqrt{p})^2 \le (1+\sqrt[4]{N})^2 < q\,. \qquad ❑$$

In order to make sure that an algorithm based on this will indeed work for every prime number, we need the following converse.

LEMMA existence of suitable points

15.2. **Lemma.** *Let N be a prime and let E be an elliptic curve over $\mathbb{Z}/N\mathbb{Z}$. Let $m = \#E(\mathbb{Z}/N\mathbb{Z})$ and let q be a prime divisor of m such that $q > (\sqrt[4]{N}+1)^2$. Then there exists a point $P \in E(\mathbb{Z}/N\mathbb{Z})$ satisfying* (15.1).

Proof. First, every point $P \in E(\mathbb{Z}/N\mathbb{Z})$ obviously satisfies $m \cdot P = O$. Since N is prime, the second condition simply says that $(m/q) \cdot P \ne O$. Now assume that no point satisfies this second condition, i.e., that $(m/q) \cdot P = O$ for all points P. We know that $E(\mathbb{Z}/N\mathbb{Z}) \cong \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/dd'\mathbb{Z}$; this then implies $dd' \mid m/q$ and so

$$m = \#E(\mathbb{Z}/N\mathbb{Z}) = d^2d' \le (dd')^2 \le (m/q)^2\,,$$

i.e., $q^2 \le m$, from which we deduce the contradiction

$$N + 6\sqrt{N} + 1 < (\sqrt[4]{N}+1)^4 < q^2 \le m \le (\sqrt{N}+1)^2 = N + 2\sqrt{N} + 1\,. \qquad ❑$$

Of course, we also need to know that a suitable curve E exists (i.e., such that $\#E(\mathbb{Z}/N\mathbb{Z})$ has a sufficiently large prime divisor). For large primes N, this follows from the existence of enough numbers with a large prime divisor in the Hasse interval (and further results saying that most of the numbers in the interval occur as orders of elliptic curves). For small primes, one can check it explicitly, but the main use case is clearly for large primes.

S. Goldwasser
* 1958

This leads to the following algorithm due to *Goldwasser* and *Kilian*.

15.3. **Algorithm.**

0. The input is a (large) natural number N that is very likely prime (in particular, $6 \perp N$).
1. We randomly choose elements a and b in $\mathbb{Z}/N\mathbb{Z}$ such that $4\,a^3 + 27\,b^2$ is invertible. Let E be the elliptic curve over $\mathbb{Z}/N\mathbb{Z}$ given by $y^2 = x^3 + a\,x + b$.
2. We compute $m = \#E(\mathbb{Z}/N\mathbb{Z})$ using the polynomial-time algorithm of Schoof-Elkies-Atkin. (If something goes wrong, then N is not prime.)
3. We factor $m = u \cdot q$ by trial division (with a reasonable bound), where u is the product of the small prime divisors found in this way, Then we check whether $(\sqrt[4]{N}+1)^2 < q \le m/2$ and q passes the Miller-Rabin test. If this is not the case, we try again with a new curve (Step 1.).
4. We randomly choose $x \in \mathbb{Z}/N\mathbb{Z}$ until the Jacobi symbol $\left(\frac{x^3+ax+b}{N}\right)$ has the value 0 or 1. Then we find $y \in \mathbb{Z}/N\mathbb{Z}$ such that $y^2 = x^3 + a\,x + b$. (If the square root algorithm fails, then we know that N is not prime.)
   Set $P = (x : y : 1) \in E(\mathbb{Z}/N\mathbb{Z})$.

5. We check that $m \cdot P = O$. If this is not the case (or an error occurs in the computation), then N is not prime.
6. If $u \cdot P = O$, then we look for another point on E (Step 4.). Otherwise, $u \cdot P = (\xi : \eta : \zeta)$ with $\zeta \neq 0$. Either ζ is not invertible; then N is not prime, or else it is, then N is prime according to Lemma 15.1, *if* q is prime.
7. To finish the proof, we apply the algorithm recursively to q (until q is small enough to be shown prime by a more direct method). If q turns out to be composite, we try again with a new curve (Step 1.).

One can show that this algorithm has an expected running time of $O((\log N)^{12})$ (under reasonable assumptions on the distribution of prime numbers in short intervals). For practical purposes, the exponent is still too large, however. The main bottleneck is the determination of $\#E(\mathbb{Z}/N\mathbb{Z})$.

There is a variant of this algorithm (due to Atkin and Morain) that essentially constructs special elliptic curves (with known endomorphism ring), for which the number m is known beforehand. This version has been implemented and can routinely check 1000-digit numbers for primality. (My experience with the Magma implementation is somewhat mixed, though.) The fastest variants of this kind of test have a heuristic running time of $O((\log N)^{4+\varepsilon})$ for arbitrary small $\varepsilon > 0$. In practice, they are at least as good (i.e., fast) as another fast primality test (that works with so-called Jacobi sums and uses quite heavy algebraic number theory; its complexity is $O((\log N)^{c \log\log\log N})$, which is slightly worse than polynomial).

It should be mentioned that the Goldwasser-Kilian test and the Atkin-Morain test have, compared to the Jacobi sum test, the advantage that they provide us with a primality *certificate* for N: Using the data E, P, m, q (and the primality certificate for q) and Lemma 15.1, one can quickly verify that N is indeed prime.

Adleman and Huang (using curves of genus 2) have constructed an algorithm whose running time is provably (not just heuristically) polynomial, but it is so far not practical. This algorithm is (like the other ones discussed in this section) probabilistic; the theoretical result that there exists a (probabilistic) polynomial-time algorithm for testing primality has been superseded by the better result of Agrawal, Kayal and Saxena.

Factorization.

To factor a number N (which is known to be composite, e.g., because it failed the Miller-Rabin test), one can proceed in the same way as in the $p - 1$ method. Instead of the multiplicative group, one uses the group of rational points on an elliptic curve.

So let E be an elliptic curve over $\mathbb{Z}/N\mathbb{Z}$ and let $P \in E(\mathbb{Z}/N\mathbb{Z})$ be a point. Let further p be a prime divisor of N. Then we obtain an elliptic curve E' over $\mathbb{F}_p$ from E (be reducing the equation mod p) and the canonical map $E(\mathbb{Z}/N\mathbb{Z}) \to E'(\mathbb{F}_p)$. Let m be the order of the image P' of P in $E'(\mathbb{F}_p)$. We assume that m is B-powersmooth. Then $L(B) \cdot P' = O$ on E'. (Recall that $L(B) = \operatorname{lcm}(1, 2, \dots, B)$.) Usually, the order of the image of P on the reductions of E modulo other prime divisors of N will not B-powersmooth, which means that $L(B) \cdot P$ is, on the one hand, not the point O (on E), but on the other hand must have the form $(\xi : \eta : \zeta)$ with ζ nonzero and not invertible (since ζ vanishes mod p). In this case, the gcd of (a representative of) ζ with N will lead to a nontrivial factor of N.

In practice, already at some earlier point during the computation we will have the situation that a division cannot be performed, since the divisor is not invertible (and also nonzero). In this case one obtains a nontrivial factor from the extended gcd computation that tries to find the inverse.

Also, it is advisable to choose the curve in such a way that it has a known point P, since computing square roots modulo N is about as hard as factoring N. Therefore, one picks, for example, $P = (1, 1)$ and then one chooses an equation of the form

$$y^2 = x^3 + Ax - A \qquad \text{or} \qquad y^2 = x^3 + Ax^2 + Bx - (A + B)\,.$$

In addition, one can work with several curves in parallel and stop as soon as one of the computations is successful.

This leads to the following algorithm.

Note that the discriminant of $y^2 = x^3 + Ax - A$ is $-2^4A^2(4A + 27)$.

15.4. **Algorithm.**

Input: N (the number to be factored, with $6 \perp N$)
B (a parameter as above), m (number of curves used)

1. For $i = 1, \ldots, m$ repeat Steps 2 to 5.
2. Choose $A \in \{1, \ldots, N - 1\}$ randomly
   and set $d_1 = \gcd(A, N)$, $d_2 = \gcd(4A + 27, N)$.
3. *(Discriminant invertible?)*
   If $d_1 > 1$, return d_1 as a factor; stop.
   If $1 < d_2 < N$, return d_2 as a factor; stop.
   If $d_2 = N$, go to Step 2.
4. Set $E\colon y^2 = x^3 + \bar{A}x - \bar{A}$ over $\mathbb{Z}/N\mathbb{Z}$ and $P = (\bar{1}, \bar{1}) \in E(\mathbb{Z}/N\mathbb{Z})$.
5. *(Computation of $L(B) \cdot P$)*
   For $p \in \{\text{primes} \le B\}$, set $P = p^{\lfloor \log_p B \rfloor} \cdot P$.
   Here we use the formulas that are valid for elliptic curves over a field. If during the course of the computation a nonzero, but not invertible element $\bar{d} \in \mathbb{Z}/N\mathbb{Z}$ shows up, return $\gcd(d, N)$ as a factor;
6. Print “No factor was found”; stop.

The efficiency of this method depends on how many B-powersmooth numbers exist near p. If we set

$$\ell(x) = e^{\sqrt{\log x \log\log x}}\,,$$

then we have the following.

15.5. **Theorem.** (Canfield, Erdős, Pomerance)[3] *The density of $\ell(x)^a$-powersmooth numbers near x is close to $\ell(x)^{-1/(2a)}$ (as x tends to infinity).*

THM
density of
powersmooth
numbers

So when we want to find prime factors up to size about M, we set $B = \ell(M)^a$ for some to-be-determined $a > 0$. Then we have to try about $\ell(M)^{1/(2a)}$ curves until we find a suitable one. For each of these curves, we need to compute the multiple $L(B) \cdot P$. This requires $O(\log L(B))$ operations (additions or doublings) on the curve (which have complexity at most $O((\log N)^2)$; we will neglect this factor). We therefore need a bound for $\log L(B)$.

15.6. Theorem. *As $B \to \infty$, we have $\log L(B) \sim B$.*

THM
growth
of $\log L(B)$

This statement is equivalent to the Prime Number Theorem, which says that the number $\pi(x)$ of prime numbers $\le x$ satisfies the asymptotic relation

$$\pi(x) \sim \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x}\,.$$

The computing time for each curve therefore is $O(B) = O(\ell(M)^a)$. In total, we obtain something of the order of $\ell(M)^{a+1/(2a)}$. This is minimized for $a = 1/\sqrt{2}$, leading to an (expected) computing time of roughly $O(\ell(M)^{\sqrt{2}})$. This shows a nice property of this method: the complexity mainly depends on the size of the prime factors we want to find. Hence it is well-suited to find smallish to medium-sized prime factors (and with some luck, what remains is prime). In the worst case, $M = \sqrt{N}$, and the computing time is about $O(\ell(N))$. Note that this is *subexponential* in $\log N$, i.e., it grows more slowly than every function $e^{c \log N} = N^c$ (with $c > 0$). In this worst case, the complexity is comparable with the quadratic sieve, which, however, is faster by a large constant factor.

Compared to other methods like the quadratic sieve, the "Elliptic Curve Method" also has the advantage to need only very little memory. On the other hand, other methods tend to be faster in practice when N is a product of two prime numbers of roughly the same size.

Here is a comparison of the various complexity classes:

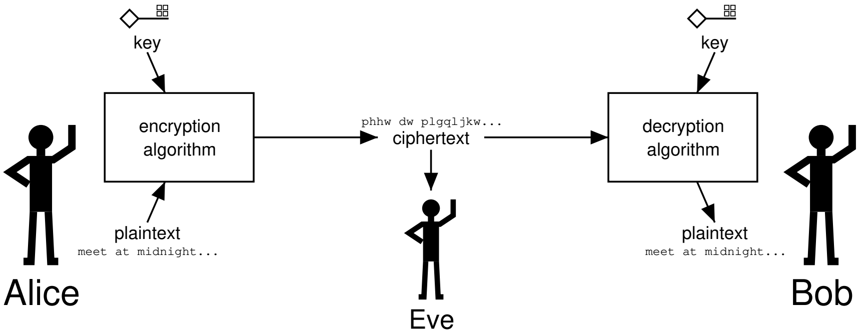
N	$\sqrt{N}$	$\sqrt[4]{N}$	$e^{\sqrt{\log N \log\log N}}$	$e^{\sqrt[3]{\log N (\log\log N)^2}}$	$(\log N)^5$	$(\log N)^{12}$
1000	10	3,2	38,6	19,2	$1{,}6 \cdot 10^4$	$1{,}2 \cdot 10^{10}$
10^6	1000	31,6	413	96,3	$5{,}0 \cdot 10^5$	$4{,}8 \cdot 10^{13}$
10^{10}	10^5	316	4910	444	$6{,}5 \cdot 10^6$	$2{,}2 \cdot 10^{16}$
10^{20}	10^{10}	10^5	$6 \cdot 10^5$	6460	$2{,}1 \cdot 10^8$	$9 \cdot 10^{19}$
10^{50}	10^{25}	$3 \cdot 10^{12}$	$1{,}4 \cdot 10^{10}$	$9 \cdot 10^5$	$2{,}0 \cdot 10^{10}$	$5 \cdot 10^{24}$
10^{100}	10^{50}	10^{25}	$2{,}3 \cdot 10^{15}$	$1{,}7 \cdot 10^8$	$6{,}4 \cdot 10^{11}$	$2{,}2 \cdot 10^{28}$
10^{200}	10^{100}	10^{50}	$1{,}2 \cdot 10^{23}$	$1{,}7 \cdot 10^{11}$	$2{,}1 \cdot 10^{13}$	$9 \cdot 10^{31}$
10^{500}	10^{250}	10^{125}	$1{,}3 \cdot 10^{39}$	$5 \cdot 10^{16}$	$2{,}0 \cdot 10^{15}$	$5 \cdot 10^{36}$
10^{1000}	10^{500}	10^{250}	10^{58}	$2{,}8 \cdot 10^{22}$	$6{,}5 \cdot 10^{16}$	$2{,}2 \cdot 10^{40}$

Factorization algorithms that are implemented in Computer Algebra Systems usually use several methods in succession. After trial division by prime numbers from a given list, one checks if the remaining factor is prime. If it is not, one can use the $p-1$ and $p+1$ methods (with not too large B). For the next step, the Elliptic Curve Method is a good choice to find moderately large factors (say, 20–30 digits). If this leaves a composite number, one uses the Multiple Polynomial Quadratic Sieve. The Number Field Sieve is not yet sufficiently efficient and robust to be used for "everyday" computations.

[3] E.R. Canfield, P. Erdős, C. Pomerance: *On a problem of Oppenheim concerning "factorisatio numerorum"*, J. Number Theory **17** (1983), no. 1, 1–28.

16. Cryptography: basics

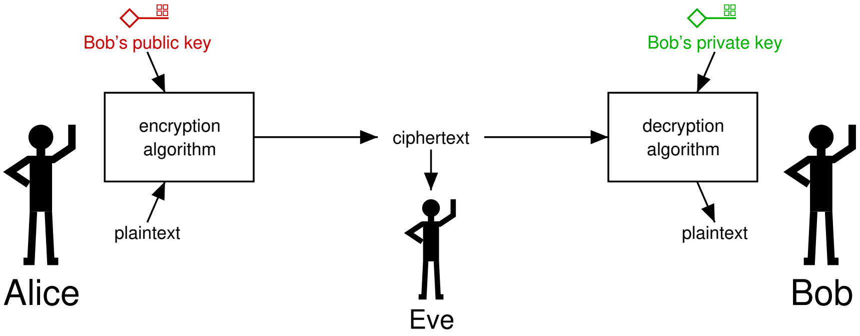
The basic task of cryptography consists in transmitting a secret message ("plaintext") securely from a sender ("Alice") to a recipient ("Bob"), even though the transmission channel is public (so that "Eve" can eavesdrop on the conversation). So the message must be encrypted (obtaining a "ciphertext") in such a way that it cannot be reconstructed by potential eavesdroppers. Classically, one uses methods that need a key that has to be kept secret for the encryption as well as for the decryption of the message. This is called a *symmetric* cryptosystem.



These methods have the advantage that they are usually very efficient, so that one is able to transmit large amounts of information quickly. The current standard (from 2001) is the *Advanced Encryption Standard* AES.

The disadvantage of symmetric methods is that Alice and Bob need to agree beforehand on a common secret key. This is difficult to arrange when they have not yet communicated, since the communication establishing the key must also be secret. This kind of situation routinely occurs for example when business transactions are to be performed via the internet. A further disadvantage of symmetric systems is that they need a separate key for each *pair* of participants. Assuming that keys are generated and distributed centrally (e.g., in a military context), this quickly becomes unmanageable as the number of participants grows.

New ideas are therefore necessary. One possible approach is to use *distinct* keys for encryption and decryption. Then the encryption key for a given participant can be publicly available; it is therefore called the participant's *public key*, while the decryption key is known only to the participant and constitutes their *private key*. Such methods are known as *asymmetric* or also as *public key cryptosystems*.



This set-up leads to stronger requirements for the encryption and decryption methods, however, since it should not be possible (at least not without undue effort) to reconstruct the plaintext from the ciphertext and the known encryption key.

Mathematically speaking, one needs something called a "one-way trapdoor function"; this is a function that can be (fairly) easily computed, but is hard to invert ("one-way"), however, inverting becomes easy when some additional information is provided ("trapdoor"). The most well-known of these methods is the RSA cryptosystem (named after the inventors Rivest, Shamir and Adleman). It is based on the empirical observation that factoring sufficiently large numbers appears to be very difficult. It works as follows.

16.1. **Example.** The RSA cryptosystem.

EX
RSA

1. Pick two random large prime numbers $p \neq q$ and set $N = pq$.
2. Choose randomly $1 < e < (p-1)(q-1)$ such that $e \perp \operatorname{lcm}(p-1, q-1)$.
3. Compute d such that $de \equiv 1 \bmod \operatorname{lcm}(p-1, q-1)$.
4. **Public key:** (N, e)
   Private key: d.
5. **Encryption:**
   $\{0, 1, \ldots, N-1\} \ni m \longmapsto c = (m^e \bmod N) \in \{0, 1, \ldots, N-1\}$.
6. **Decryption:**
   $\{0, 1, \ldots, N-1\} \ni c \longmapsto m = (c^d \bmod N) \in \{0, 1, \ldots, N-1\}$. ♣

Fermat's Little Theorem makes this method work:

$$m^{de} = m \cdot (m^{p-1})^a \equiv m \bmod p$$

and in the same way

$$m^{de} = m \cdot (m^{q-1})^b \equiv m \bmod q\,,$$

where we set $de = 1 + a(p-1) = 1 + b(q-1)$. This implies that $m^{de} \equiv m \bmod N$.

The security of the RSA system comes down to the difficulty of computing d from e and N when the prime divisors p and q are not known. This is about as hard as finding these primes, as is shown by the following result.

16.2. **Lemma.** *Let $p, q > 2$ be two prime numbers, $N = pq$, and $d, e \in \mathbb{Z}$ such that $de \equiv 1 \bmod \operatorname{lcm}(p-1, q-1)$. We write $de - 1 = 2^t u$ with u odd. Then for at least half the integers $1 \leq a < N$ such that $a \perp N$, one of the numbers*

LEMMA
RSA and factorization

$$\gcd(a^u - 1, N)\,, \quad \gcd(a^{2u} - 1, N)\,, \quad \ldots\,, \quad \gcd(a^{2^{t-1}u} - 1, N)$$

is a nontrivial divisor of N.

Proof. By definition, $2^t u$ is a multiple of $p-1$ and of $q-1$, and u is odd and therefore not a multiple of $p-1$. This implies that

$$e_p := \min\{e \geq 0 : p - 1 \mid 2^e u\} \in \{1, 2, \ldots, t\}$$

and similarly for the analogously defined e_q. Then $a^{2^{e_p} u} \equiv 1 \bmod p$ for all $a \perp N$, which shows that $a^{2^{e_p - 1} u} \equiv \pm 1 \bmod p$, and both possibilities occur the same number of times. The analogous statement holds for q and e_q.

If $e_p > e_q$, then half of all a satisfy

$$a^{2^{e_p-1}u} \equiv -1 \bmod p \quad \text{and} \quad a^{2^{e_p-1}u} \equiv 1 \bmod q \quad \Longrightarrow \quad \gcd(a^{2^{e_p-1}u} - 1, N) = q\,.$$

In a similar way, we obtain the divisor p as a gcd in half of the cases, if $e_p < e_q$.

In the remaining case $e_p = e_q =: e$, we have that

$$a^{2^{e-1}u} \equiv \varepsilon_p \bmod p \quad \text{and} \quad a^{2^{e-1}u} \equiv \varepsilon_q \bmod q$$

for a quarter of all a, for each combination of signs $\varepsilon_p, \varepsilon_q = \pm 1$. (By the Chinese Remainder Theorem, the residue classes of a modulo p and modulo q are independent). We obtain different signs for half of the residue classes; then $\gcd(a^{2^{e-1}u} - 1, N)$ is either p or q. ❑

So when we know both the encryption exponent e and the decryption exponent d, then we will find with on average at most two tries the two prime factors p and q of N via a computation whose complexity is comparable with encryption/decryption.

We have seen that by now there exist factorization algorithms of subexponential complexity. This means that one has to use fairly long keys to obtain a secure method, which has a negative effect on the efficiency. (And there is a fast factorization algorithm using quantum computers.)

Another method is based on the difficulty to compute *discrete logarithms* in multiplicative groups.

16.3. **Definition.** Let $G = \langle g \rangle$ be a finite cyclic group (written multiplicatively) with given generator g. Then every element $h \in G$ can be written in the form $h = g^a$, and we call a (which is uniquely determined modulo $\#G$) the *discrete logarithm* of h to base g. ♢

DEF
discrete logarithm

One application of this is the *Diffie-Hellman key exchange*. This is not used to encrypt a message, but to generate a secret that is shared between the participants. This secret information can then for example be used as the key for a symmetric cryptosystem.

16.4. **Example.** The Diffie-Hellman key exchange.

EX
Diffie-Hellman

1. The participants agree to use a finite cyclic group G with generator g.
   The original method uses $G = \mathbb{F}_p^\times$ with a large prime p.
2. Alice chooses a number a randomly and computes $A = g^a$.
   Bob chooses a number b randomly and computes $B = g^b$.
3. Alice sends A to Bob. Bob sends B to Alice.
4. Alice computes $s = B^a$. Bob computes $s = A^b$. ♣

Since $A^b = (g^a)^b = g^{ab} = g^{ba} = (g^b)^a = B^a$, both indeed compute the same element $s \in G$. To determine the secret s from the public data G, g, A, B, one has to solve the so-called Diffie-Hellman problem. This is certainly possible when one can compute discrete logarithms in G, since then we obtain for example a as the logarithm of A and so can compute $s = B^a$ like Bob does. Conjecturally, both problems (Diffie-Hellman and discrete logarithms) are of comparable difficulty.

One can also use the underlying idea to directly encrypt messages.

16.5. **Example.** The cryptosystem after ElGamal.

EX
ElGamal

1. Alice and Bob agree on a group G of order n with generator g.
2. Bob chooses a random number $b \in \mathbb{Z}/n\mathbb{Z}$.
3. **Private key:** b
   Public key: $h = g^b$.

4. **Encryption:**
   Alice chooses $a \in \mathbb{Z}/n\mathbb{Z}$ randomly and computes from the plaintext $m \in G$ the pair $(r, s) = (g^a, h^a \cdot m)$.
5. **Decryption:** Bob computes $m = r^{-b} \cdot s$. ♣

Originally, these methods were proposed for the multiplicative group $G = \mathbb{F}_p^\times$. However, over time, algorithms have been developed that compute discrete logarithms in multiplicative groups of finite fields with a complexity comparable to the best factorization algorithms. The level of security for given key length is therefore comparable with that of RSA.

There are "generic" methods for the computation of discrete logarithms in arbitrary cyclic groups; they are based on variants of the birthday paradox and have a complexity of $O(\sqrt{n})$ group operations, where n is the group order. If n is not prime, one can reduce the problem to smaller groups whose order is a prime divisor of n. Combined with a generic method for these groups, this leads to a complexity of $O(\sqrt{p})$ group operations, where p is the largest prime divisor of n. Therefore, one should choose group orders that are prime (or "almost" prime, i.e., a small factor times a prime). A more detailed discussion of algorithms for the discrete logarithm problem can be found in the small print below.

Before we discuss how one can use elliptic curves with profit in this context, I would like to mention some algorithms for discrete logarithms.

We assume we are given a (finite) cyclic group G with generator g and known order $n = \#G$. We are also given an element $h \in G$, and we want to find $a \in \mathbb{Z}/n\mathbb{Z}$ such that $h = g^a$.

ALGO
Brute force

```
1. Set $x := 1_G$.
2. For $a = 0, 1, \ldots, n-1$, execute Steps 3 and 4.
3. If $h = x$, then return $a$; stop.
4. Set $x := x \cdot g$.
```

It is clear that the expected running time (in terms of the number of performed operations in G) is of the order of n and hence exponential in the size $O(\log n)$ of the input.

It is also clear that every other (reasonable) algorithm will be better than this one.

One possible improvement is to not compare one element with all elements of G, but to look for a common element in two roughly equal-sized sets. This idea is related to the birthday paradox; it leads to the following algorithm.

ALGO
baby-step-giant-step

```
1. Set $m := \lceil\sqrt{n}\rceil$ and $\gamma := g^m$.
2. Compute $\gamma^0, \gamma^1, \ldots, \gamma^{m-1}$ and store the pairs $(j, \gamma^j)$ in a table $T$.
3. For $r = 0, 1, \ldots, m-1$, execute Steps 4 and 5.
4. Compute $k := hg^{-r}$ and check, if there is an entry $(j, k)$ in $T$.
5. If the entry exists, then return $jm + r$; stop.
```

This approach is based on the following consideration. We have $n \leq m^2$, so $a \leq n - 1 < m^2$; this allows us to write

$$a = qm + r$$

with $q \leq m - 1$ and $0 \leq r \leq m - 1$. The equality $h = g^a$ holds if and only if

$$hg^{-r} = (g^m)^q \,.$$

We therefore first compute all possible values of the right hand side (in Step 2) and then all possible values of the left hand side (in Step 4) until we find a common value. The table T has to be organized in such a way that one can easily find an entry from its second components. This can be achieved for example using hashtables.

The complexity is $O(\sqrt{n})$ operations in G. This is still exponential in $\log n$, but already quite a bit better than the simple brute-force method. The disadvantage is that this method also needs $O(\sqrt{n})$ of memory. This can be problematic when n is large.

The next method uses a similar idea, but does not need much memory. We need a function

$$f = (f_1, f_2)\colon G \to \mathbb{Z} \times \mathbb{Z}\,,$$

that is “sufficiently random”. For example, we can extract a few bits from the internal representation of the group elements and associate to the various bit patterns previously chosen random numbers as values of f_1 and f_2. Four of five bits are usually sufficient. We then define (depending on the input G, g, h)

$$F\colon G \longrightarrow G\,, \qquad z \longmapsto z \cdot g^{f_1(z)} \cdot h^{f_2(z)}\,.$$

If $z = g^a \cdot h^b$, then $F(z) = g^{a+f_1(z)} \cdot h^{b+f_2(z)}$. We additionally choose a (relatively large) number M.

ALGO
Pollard lambda

```
1. Pick $x_0, y_0, x'_0, y'_0 \in \mathbb{Z}$ randomly and set $z_0 := g^{x_0} \cdot h^{y_0}$ and $z'_0 := g^{x'_0} \cdot h^{y'_0}$.
   Initialize am empty table $T$.
2. For $m = 1, 2, \ldots$, execute Steps 3–6.
3. Set $z_m := F(z_{m-1})$, $(x_m, y_m) := (x_{m-1}, y_{m-1}) + f(z_{m-1})$.
4. If $T$ contains an entry $(x, y, z_m)$ and $y - y_m$ is invertible modulo $n$, then compute a solution $a$ of
       $a(y - y_m) \equiv x_m - x \bmod n$
   and return $a$; stop.
   If $y - y_m$ is not invertible modulo $n$, then go to Step 1.
5. Set $z'_m := F(z'_{m-1})$, $(x'_m, y'_m) := (x'_{m-1}, y'_{m-1}) + f(z'_{m-1})$.
6. If $m$ is divisible by $M$, then store $(x'_m, y'_m, z'_m)$ in $T$.
```

This method computes two sequences $z_m = g^{x_m} \cdot h^{y_m}$ and $z'_m = g^{x'_m} \cdot h^{y'_m}$ in G and tries to find a collision $z_m = z'_{m'}$. In this case, we have the relation

$$g^{x_m} \cdot h^{y_m} = g^{x'_{m'}} \cdot h^{y'_{m'}} \qquad \Longrightarrow \qquad g^{a(y'_{m'} - y_m)} = h^{y'_{m'} - y_m} = g^{x_m - x'_{m'}}\,,$$

and if $y'_{m'} - y_m$ is invertible modulo the group order n, then we can solve this for the discrete logarithm a. If we cannot solve the congruence uniquely, we can pick new initial values (and possible change the function f). If $y'_{m'} \not\equiv y_m \bmod n$, then we obtain at least partial information on a that we can use later. In cryptographic applications, the group order n is usually a prime number, however, so that this case cannot occur.

This algorithm is also called the “method of tame and wild kangaroos”. The tame kangaroo jumps through the group (the sequence (z'_m)) and digs a hole after every Mth jump. The wild kangaroo also jumps through G (the sequence (z_m)). Eventually, it will hit the trail of the tame kangaroo and then it will be caught in a hole after at most $M - 1$ further jumps.

Similarly as for the Pollard rho factorization method, one can show that (assuming f is chosen randomly) one obtains a collision after expected $O(\sqrt{n})$ steps. The time complexity is therefore $O(\sqrt{n} + M)$, and one needs $O(\sqrt{n}/M)$ of memory. So one can keep the memory almost constant without paying too much in time. In particular, one can choose M according to the available memory.

Pohlig-Hellman reduction.

If the group order n is not prime and its prime factorization is known, then the computation of discrete logarithms in G can be reduced to the computations of discrete logarithms in groups of order p, where p runs through the prime divisors of n. This approach goes back to Pohlig and Hellman[4].

[4] G.C. Pohlig, M.E. Hellman: *An improved algorithm for computing logarithms over* GF(p) *and its cryptographic significance,* IEEE Trans. Information Theory **IT-24**, 106–110 (1978).

Let $n = p_1^{e_1} \cdots p_k^{e_k}$. In a first step, we reduce the problem to the computation of discrete logarithms in subgroups of order $p_j^{e_j}$ (for $j = 1, \ldots, k$). Note that G has a unique such subgroup for each j, namely

$$G_j = \{\gamma \in G \mid \gamma^{p_j^{e_j}} = 1_G\} = \{\gamma^{c_j} \mid \gamma \in G\}\,,$$

where $c_j = n/p_j^{e_j}$. Then $h^{c_j}, g^{c_j} \in G_j$ and $h^{c_j} = (g^{c_j})^a$. If we compute the discrete logarithm of h^{c_j} to base g^{c_j} in G_j, then this gives us $a \bmod p_j^{e_j}$. Using the Chinese Remainder Theorem, we then reconstruct a.

Now we assume that G has prime power order $n = p^e$. We first determine $a \bmod p$. Similarly as above, we note that

$$G' = \{\gamma^{p^{e-1}} \mid \gamma \in G\}$$

is the (unique) subgroup of order p of G. We compute the discrete logarithm of $h^{p^{e-1}}$ to base $g^{p^{e-1}}$ in G'; this gives $a \bmod p$. Say, $a \equiv a_0 \bmod p$. Then hg^{-a_0} is contained in the subgroup

$$G'' = \{\gamma^p \mid \gamma \in G\} = \{\gamma \in G \mid \gamma^{p^{e-1}} = 1_G\}$$

of order p^{e-1} that is generated by g^p. We recursively compute the discrete logarithm a' of hg^{-a_0} to base g^p. Then

$$hg^{-a_0} = g^{a'p} \qquad \Longrightarrow \qquad h = g^{a_0 + a'p}\,,$$

and so $a = a_0 + a'p$.

If one combines the Pohlig-Hellman reduction with Pollard lambda or Baby-step-giant-step, then the complexity is reduced to essentially $O(\sqrt{p}$, where p is the largest prime divisor of $n = \#G$.

For cryptographic applications, one clearly wants that discrete logarithms are hard to find. This is why one uses groups for this whose order is a prime (or a prime times a small factor).

The algorithms described so far are *generic*, i.e., they can be applied to arbitrary groups G (as long as we can compute products and inverses and compare elements in G). Now I want to describe a method that is specific for $G = \mathbb{F}_p^\times$.

We choose a bound B and set $F_B = \{p \mid p \text{ prime}, p \leq B\}$; this set F_B is again the *factor basis*. Here g is a primitive root mod p.

ALGO
index calculus

1. Initialize an empty list L.
2. Repeat Steps 3 and 4 until $\#L \geq \#F_B + 10$.
3. Pick a random $x \in \{1, \ldots, p-2\}$ and compute $y = g^x \bmod p$.
4. If y is B-smooth, then write $y = \prod_{q \in F_B} q^{e_q}$ and store $(x, (e_q)_{q \in F_B})$ in L.
5. Solve the following linear system of equations over $\mathbb{Z}/(p-1)\mathbb{Z}$ in the unknowns a_q, $q \in F_B$: For each entry $(x, (e_q)_{q \in F_B})$ in L, we have the equation
$$\sum_{q \in F_B} e_q a_q = x\,.$$
6. (*Here we have* $q \equiv g^{a_q} \bmod p$ *for all* $q \in F_B$)
   Repeat Steps 7 and 8 until we are successful.
7. Pick a random $x \in \{0, \ldots, p-2\}$ and compute $y = g^x h \bmod p$.
8. If y is B-smooth, write $y = \prod_{q \in F_B} q^{e_q}$ and return $\sum_{q \in F_B} e_q a_q - x$.

Similarly to the Quadratic Sieve, we first generate relations between g and the primes in the factor basis. Then these relations are used to determine the discrete logarithms of these primes. Finally, this information is used to solve the original problem. If one needs to compute many discrete logarithms in the same group $\mathbb{F}_p^\times$, then one can of course store the result of Step 5 and then immediately start with Step 6 for each new h.

The complexity analysis is again based on Theorem 15.5 by Canfield, Erdős and Pomerance. If one chooses B optimally, one gets a running time of $O(e^{c\sqrt{\log p \log\log p}})$, comparable with the Quadratic Sieve. It is also possible to adapt the number field sieve for the computation of discrete logarithms, which again leads to a complexity of $O(e^{c\sqrt[3]{\log x(\log\log x)^2}})$.

17. Cryptography: elliptic curves

In a similar way as using elliptic curves leads to a much more flexible version of the $(p-1)$ factoring method by replacing the multiplicative group $\mathbb{F}_p^\times$ with a group $E(\mathbb{F}_p)$, we can replace the multiplicative group by the group of $\mathbb{F}_q$-rational points on an elliptic curve in cryptographic applications. The methods remain the same (as described in the previous section for general cyclic groups). The only difference is that we write the group additively. This gives the following versions.

We first need to fix an elliptic curve E over a finite field $\mathbb{F}_q$, together with a point $P \in E(\mathbb{F}_q)$ whose order is a sufficiently large prime number n. We work with the group $G = \langle P \rangle$. (We take a group of prime order, since otherwise the discrete logarithm problem can be reduced to smaller groups.)

17.1. **Example.** Diffie-Hellman key exchange using elliptic curves.

EX
Diffie-Hellman
with ell. curve

(1) Alice chooses a random number a and computes $A = a \cdot P$.
Bob chooses a random number b and computes $B = b \cdot P$.

(2) Alice sends A to Bob. Bob sends B to Alice.

(3) Alice computes $S = a \cdot B$. Bob computes $S = b \cdot A$. ♣

17.2. **Example.** ElGamal encryption using elliptic curves.

EX
ElGamal
with ell. curve

(1) Bob chooses a random number $b \in \mathbb{Z}/n\mathbb{Z}$.

(2) **Private key:** b,
Public key: $B = b \cdot P$.

(3) **Encryption:**
Alice chooses a random $a \in \mathbb{Z}/n\mathbb{Z}$
and computes from the plaintext $M \in G$ the pair $(R, S) = (a \cdot P, a \cdot B + M)$.

(4) **Decryption:** Bob computes $M = S - b \cdot R$. ♣

There are further protocols, for example for digital signatures or authentication.

Why is it advantageous to use elliptic curves instead of multiplicative groups? In multiplicative groups, discrete logarithms can be computed in subexponential time (comparable to factorization). This means in practice that one needs fairly large key lengths (several 1000 bits) to obtain a reasonable level of security. This has obvious implications for the efficiency of encryption and decryption, leading to significantly slower operation than symmetric methods. In addition, it is hard to implement such a protocol on hardware with very limited resources (think smartcards).

The essential advantage of elliptic curves then is that (at least so far) no algorithm for the computation of discrete logarithms is known that would apply to arbitrary elliptic curves and is faster than the generic algorithms (of complexity $O(\sqrt{n})$). This means that one can get by with using significantly shorter keys (a few 100 bits). So on the one hand, encryption and decryption are faster than with comparably secure protocols based on multiplicative groups (even though the individual operations in the group require more effort), and on the other hand, one needs a fairly small amount of memory, making these methods well-suited for smartcards and similar devices. It is well possible that you carry around one or several elliptic curves in your wallet!

There are, however, avenues for attack in certain situations.

G. Frey
* 1944
Foto © MFO

One attack, the so-called Frey-Rück attack[5] is based on the *Tate pairing* (which is related with the Weil pairing). Let E be an elliptic curve over a finite field $\mathbb{F}_q$ and n a number coprime with q. The Tate pairing is a map

$$\langle \cdot, \cdot \rangle_{\text{Tate}} \colon E(\mathbb{F}_q)/nE(\mathbb{F}_q) \times E(\mathbb{F}_q)[n] \longrightarrow \mathbb{F}_q^\times/\mathbb{F}_q^{\times n} .$$

To compute $\langle P + nE(\mathbb{F}_q), Q\rangle_{\text{Tate}}$, let $F_Q \in \mathbb{F}_q(E)$ be a rational function on E with an n-fold zero in Q and an n-fold pole in O. Write $P = P_1 - P_2$ with $\{P_1, P_2\} \cap \{Q, O\} = \emptyset$. Then

$$\langle P, Q\rangle_{\text{Tate}} = \langle P + nE(\mathbb{F}_q), Q\rangle_{\text{Tate}} = \frac{F_Q(P_1)}{F_Q(P_2)} \cdot \mathbb{F}_q^{\times n} .$$

One can show that this definition does not depend on the choice of F_Q (this is easy since the possible choices differ only by scaling), the choice of the representative P or the choice of the representation of P as a difference. The Tate pairing is bilinear (in the same sense as for the Weil pairing). If $q \equiv 1 \bmod n$, then it is also non-degenerate. (Otherwise, when $q \not\equiv 1 \bmod n$, then $\mathbb{F}_q^\times/\mathbb{F}_q^{\times n}$ has order less than n, which forces the pairing to be degenerate.)

We need a Lemma.

LEMMA
field of definition of n-torsion points

17.3. **Lemma.** *Let E be an elliptic curve over $\mathbb{F}_q$ and $P \in E(\mathbb{F}_q)$ a point of prime order $n \perp q$. Let $l \geq 1$ be the smallest number such that $q^l \equiv 1 \bmod n$. If $l > 1$, then $E[n] \subset E(\mathbb{F}_{q^l})$.*

Proof. Since $n \perp q$, we have by Theorem 12.1 (1) that $E[n] = E(\bar{\mathbb{F}}_q)[n] \cong \mathbb{Z}/n\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$. We can extend P to a basis (P, Q) of $E[n]$. Let $M \in \mathrm{GL}(2, \mathbb{Z}/n\mathbb{Z})$ be the matrix of $\phi|_{E[n]}$ with respect to this basis, where ϕ is the Frobenius endomorphism of E over $\mathbb{F}_q$. Since $P \in E(\mathbb{F}_q)$, we have $\phi(P) = P$, and so M has the form

$$M = \begin{pmatrix} 1 & a \\ 0 & b \end{pmatrix} .$$

Now $\zeta = e_n(P, Q)$ is a primitive nth root of unity in $\bar{\mathbb{F}}_q$ (by Theorem 12.2 (3), e_n is non-degenerate). The compatibility of e_n with the action of the absolute Galois group of $\mathbb{F}_q$ (Theorem 12.2 (4)) implies that

$$\zeta^b = e_n(P, aP + bQ) = e_n(\phi(P), \phi(Q)) = \phi(\zeta) = \zeta^q ,$$

which shows that $b = q$ in $\mathbb{Z}/n\mathbb{Z}$. The matrix of $\phi^l|_{E[n]}$ then is

$$M^l = \begin{pmatrix} 1 & a' \\ 0 & q^l \end{pmatrix} = \begin{pmatrix} 1 & a' \\ 0 & 1 \end{pmatrix} ,$$

because $q^l = 1$ in $\mathbb{Z}/n\mathbb{Z}$ according to the definition of l. If $l > 1$, then $b = q \neq 1$ in $\mathbb{Z}/n\mathbb{Z}$, hence M has the two distinct eigenvalues 1 and q and is therefore diagonalizable. Then M^l is diagonalizable as well, which implies that $a' = 0$, so ϕ^l is the identity on $E[n]$. But this precisely means that the elements of $E[n]$ are in $E(\mathbb{F}_{q^l})$. ❑

[5] G. Frey, H.-G. Rück: *A remark concerning m-divisibility and the discrete logarithm in the divisor class group of curves*, Math. Comp. **62**, 865–874 (1994)

So let E be an elliptic curve over $\mathbb{F}_q$ and let $P \in E(\mathbb{F}_q)$ be a point of prime order n such that $n \perp q$. Let l be as in the lemma above. If $l > 1$, then the lemma shows that there is a point $P' \in E(\mathbb{F}_{q^l})[n]$ that is not in $\langle P \rangle$. Then

$$\langle P, P' \rangle_{\text{Tate}} \neq 1\,.$$

There is an isomorphism

$$\alpha\colon \mathbb{F}_{q^l}^\times / \mathbb{F}_{q^l}^{\times n} \longrightarrow \mu_n(\mathbb{F}_{q^l})\,, \qquad a \cdot \mathbb{F}_{q^l}^{\times n} \longmapsto a^{(q^l-1)/n}\,.$$

To compute the discrete logarithm of $Q \in \langle P \rangle$, we determine

$$r = \alpha\bigl(\langle P, P' \rangle_{\text{Tate}}\bigr) \quad \text{und} \quad s = \alpha\bigl(\langle Q, P' \rangle_{\text{Tate}}\bigr)\,.$$

From $Q = aP$ and the bilinearity of the Tate pairing we deduce $s = r^a$. Therefore the determination of a corresponds to the computation of a discrete logarithm in (the subgroup of order n of) $\mathbb{F}_{q^l}^\times$. If l is not too large, then the available subexponential algorithms are faster than generic algorithms for $\langle P \rangle$. In practice one should choose E and P so that $l > 20$.

If $l = 1$, then we can work directly in $E(\mathbb{F}_q)$. In this case (assuming $n^2 \nmid \#E(\mathbb{F}_q)$), $\langle P, P \rangle_{\text{Tate}}$ is nontrivial, and we can proceed as above, but with $P' = P$. This reduces the problem to a discrete logarithm in $\mathbb{F}_q^\times$ that can be computed much more easily than with generic methods. So the case $q \equiv 1 \bmod n$ must be avoided.

As a reaction to this, it was suggested to use curves E over $\mathbb{F}_p$ such that $\#E(\mathbb{F}_p) = p$, as they are immune to this kind of attack. However, it soon transpired that one can compute discrete logarithms on such curves even more easily. To do this, one chooses an elliptic curve $\tilde{E}$ over the field $\mathbb{Q}_p$ of p-adic numbers whose equation reduces mod p to that of E. By Hensel's Lemma, one can lift the points P and Q to points $\tilde{P}, \tilde{Q} \in \tilde{E}(\mathbb{Q}_p)$. The points $p\tilde{P}$ and $p\tilde{Q}$ lie in the "kernel of reduction", this is the subgroup $\tilde{E}_1(\mathbb{Q}_p)$ of $\tilde{E}(\mathbb{Q}_p)$ whose elements are the points whose reduction mod p is the origin $O \in E(\mathbb{F}_p)$. These are exactly $O \in \tilde{E}(\mathbb{Q}_p)$ and the points (ξ, η) with $v_p(\xi/\eta) > 0$ and $v_p(\xi) < 0$. We use $\tilde{E}_2(\mathbb{Q}_p)$ to denote the subgroup of points such that $v_p(\xi/\eta) \geq 2$ (together with O). Then there are isomorphisms

$$\begin{tikzcd}
E(\mathbb{F}_p) \arrow[r] & \tilde{E}_1(\mathbb{Q}_p)/\tilde{E}_2(\mathbb{Q}_p) \arrow[r] & \mathbb{Z}/p\mathbb{Z} \\
Q \arrow[r, mapsto] & p\tilde{Q} & \\
& R \arrow[r, mapsto] & \frac{x}{py}(R) \bmod p
\end{tikzcd}$$

if $p\tilde{P} \notin \tilde{E}_2(\mathbb{Q}_p)$. If this condition is not satisfied, one picks a different curve $\tilde{E}$.

The computation of discrete logarithms in $E(\mathbb{F}_p)$ thus is reduced to the computation of discrete logarithms in the *additive* group $\mathbb{Z}/p\mathbb{Z}$. This can be done quite trivially via the Extended Euclidean Algorithm.

There are further possibilities for attacks when the field $\mathbb{F}_q$ has order $q = p^m$ with a composite number m. Therefore it is recommended to either use a curve over a prime field $\mathbb{F}_p$, or else a curve over a field $\mathbb{F}_{2^p}$ with p prime. In any case one has to make sure that none of the attacks described above can be applied.

To compute the order $\#E(\mathbb{F}_q)$ one uses (for $q = p$) the algorithm of Schoof-Elkies-Atkin; for $q = 2^p$ there exists a very efficient algorithm due to Satoh. Alternatively, one can fix the group order beforehand and then construct curves with the given order (this is similar to what is done in the Atkin-Morain primality proof). This is done by using curves with complex multiplication that are defined over a suitable algebraic number fields, which are then reduced modulo a suitable prime number.

18. The rational torsion subgroup

In the remaining part of the course we will study the group $E(\mathbb{Q})$ of rational points on an elliptic curve E over $\mathbb{Q}$. As a first step, we look in more detail at the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$. For the following, we fix an elliptic curve

$$E\colon y^2 = x^3 + ax + b$$

given by a short Weierstrass equation with coefficients $a, b \in \mathbb{Z}$. Our first goal is to show that every nontrivial (i.e., $\neq O$) point of finite order in $E(\mathbb{Q})$ must have *integral* coordinates.

If $P = (\xi, \eta) \in E(\mathbb{Q})$ is a point such that ξ or η is not integral, then there is a prime number p such that $v_p(\xi) < 0$ or $v_p(\eta) < 0$ (i.e., p divides the denominator of one of the coordinates). If $v_p(\xi) < 0$, then the term ξ^3 on the right hand side $\xi^3 + a\xi + b$ of the equation of E evaluated at P is that with the smallest p-adic valuation; this implies that

$$2v_p(\eta) = v_p(\eta^2) = v_p(\xi^3 + a\xi + b) = 3v_p(\xi)\,,$$

and we see that *both* valuations are negative and that there is some $e \geq 1$ such that $v_p(\xi) = -2e$ and $v_p(\eta) = -3e$. If $v_p(\eta) < 0$, then the valuation of the right hand side must also be negative, which again implies $v_p(\xi) < 0$.

This means that the point has coordinates that are “large” in the p-adic metric (note that $|\xi|_p = p^{-v_p(\xi)}$, so $|\xi|_p \geq p^2$, $|\eta|_p \geq p^3$) and is therefore “close” to the point O at infinity. This prompts us to consider the equation of E on an affine part that contains O. So we restrict to $y \neq 0$ and dehomogenize the projective equation by setting $y = 1$. We will use coordinates $z = x/y$ and $w = 1/y$; the corresponding projective point is $(z : 1 : w)$. In these coordinates, the equation of E reads

$$E\colon w = z^3 + aw^2z + bw^3\,,$$

and the point O has coordinates $(z, w) = (0, 0)$. If P is a point whose (standard) coordinates x and y have denominator divisible by p, we then get that

$$v_p(z(P)) = v_p(x(P)/y(P)) = v_p(x(P)) - v_p(y(P)) = -2e - (-3e) = e$$

and $v_p(w(P)) = v_p(1/y(P)) = 3e$ where e is as above. Conversely, if a point P satisfies this, then the equation implies that $v_p(w(P)) = 3v_p(z(P))$, which leads to

$$\begin{aligned} v_p(y(P)) &= -v_p(w(P)) = -3v_p(z(P)) \qquad \text{and} \\ v_p(x(P)) &= v_p(z(P)) + v_p(y(P)) = -2v_p(z(P))\,. \end{aligned}$$

So the points of interest are exactly those for which (the numerators of) $z(P)$ and $w(P)$ are divisible by p.

18.1. Definition. Let $E\colon y^2 = x^3 + ax + b$ be as above and let p be a prime. For $e \geq 1$ we define

DEF
kernel of
reduction

$$\begin{aligned} E_p^{(e)}(\mathbb{Q}) &= \{P \in E(\mathbb{Q}) \mid v_p(z(P)) \geq e, v_p(w(P)) \geq 3e\} \\ &= \{P \in E(\mathbb{Q}) \mid v_p(x(P)) \leq -2e, v_p(y(P)) \leq -3e\} \cup \{O\} \end{aligned}$$

and call $E_p^{(e)}(\mathbb{Q})$ the *eth kernel of reduction modulo p* of $E(\mathbb{Q})$. $E_p^{(1)}(\mathbb{Q})$ is also simply called the *kernel of reduction modulo p* of $E(\mathbb{Q})$. ♢

The reason for this name will become clear later.

We first prove a lemma. In the following, when we write $\alpha \equiv \beta \bmod p^e$ for rational numbers α and β, we mean that $v_p(\alpha), v_p(\beta) \ge 0$ and $v_p(\alpha-\beta) \ge p^e$. (I.e., p does not divide the denominator of α or β, and p^e divides the numerator of $\alpha - \beta$.)

18.2. Lemma. *Let $P = (\zeta : 1 : \omega)$ and $P' = (\zeta' : 1 : \omega')$ be two points in $E_p^{(e)}(\mathbb{Q})$.*

LEMMA points in $E_p^{(e)}(\mathbb{Q})$

(1) $\zeta' = \zeta \implies P' = P$.

(2) *The line through P and P' (the tangent to E in P when $P = P'$) has an equation of the form $w = sz + t$ with $v_p(s) \ge 2e$, $v_p(t) \ge 3e$.*

Proof. We take the difference of the two equalities

$$\omega' = \zeta'^3 + a\omega'^2\zeta' + b\omega'^3 \qquad \text{and} \qquad \omega = \zeta^3 + a\omega^2\zeta + b\omega^3$$

and rearrange terms. This gives

$$(\omega' - \omega)\bigl(1 - a(\omega' + \omega)\zeta - b(\omega'^2 + \omega'\omega + \omega^2)\bigr) = (\zeta' - \zeta)(\zeta'^2 + \zeta'\zeta + \zeta^2 + a\omega'^2)\,.$$

(1) If $\zeta' = \zeta$, then the right hand side vanishes. Since the second factor on the left has p-adic valuation 0 and is therefore in particular nonzero, it follows that $\omega' = \omega$.

(2) First assume $P' \ne P$. Then by (1), $\zeta' \ne \zeta$, so $s = (\omega' - \omega)/(\zeta' - \zeta)$. This gives

$$s\underbrace{\bigl(1 - a(\omega' + \omega)\zeta - b(\omega'^2 + \omega'\omega + \omega^2)\bigr)}_{v_p=0} = \underbrace{\zeta'^2 + \zeta'\zeta + \zeta^2 + a\omega'^2}_{v_p \ge 2e},$$

hence $v_p(s) \ge 2e$ and $v_p(t) = v_p(\omega - s\zeta) \ge 3e$.

When $P' = P$, we obtain by implicit differentiation or passing to the limit in the expression above

$$s(1 - 2a\omega\zeta - 3b\omega^2) = 3\zeta^2 + a\omega^2\,,$$

and we can conclude in the same way. ❑

The following theorem summarizes the most important properties of $E_p^{(e)}(\mathbb{Q})$.

18.3. Theorem. *Let $E\colon y^2 = x^3 + ax + b$ be an elliptic curve over $\mathbb{Q}$ with $a, b \in \mathbb{Z}$, let p be a prime and let $e \in \mathbb{Z}_{\ge 1}$.*

THM kernel of reduction

(1) *$E_p^{(e)}(\mathbb{Q})$ is a subgroup of $E(\mathbb{Q})$.*

(2) *If $P \in E_p^{(1)}(\mathbb{Q})$ has $v_p(z(P)) = e$ and $0 \ne m \in \mathbb{Z}$, then*

$$v_p(z(mP)) = e + v_p(m)\,.$$

(3) *$E_p^{(1)}(\mathbb{Q})_{\text{tors}} = \{O\}$.*

Proof.

(1) It is clear that $O \in E_p^{(e)}(\mathbb{Q})$. Since $-(z : 1 : w) = (-z : 1 : -w)$m $E_p^{(e)}(\mathbb{Q})$ is closed under negation. Let now $P = (\zeta : 1 : \omega)$ and $P' = (\zeta' : 1 : \omega')$ be points in $E_p^{(e)}(\mathbb{Q})$. Then $P + P' = -P''$, where P'' is the third point of intersection with E of the line through P and P'. According to Lemma 18.2, this line has the form $w = sz + t$ with $v_p(s) \ge 2e$, $v_p(t) \ge 3e$. Substituting this into the equation relating w and z leads to

$$(1 + as^2 + bs^3)z^3 + (2ast + 3bs^2t)z^2 + \ldots = 0\,.$$

Setting $\zeta'' = z(P'')$ and $\omega'' = w(P'')$, we obtain

$$\zeta + \zeta' + \zeta'' = -\frac{2ast + 3bs^2t}{1 + as^2 + bs^3}\,,$$

so

$$\begin{aligned} v_p(\zeta'') &= v_p\Big(-\frac{2ast + 3bs^2t}{1 + as^2 + bs^3} - \zeta - \zeta'\Big) \\ &\ge \min\big\{v_p(2ast + 3bs^2t) - v_p(1 + as^2 + bs^3), v_p(\zeta), v_p(\zeta')\big\} \\ &\ge \min\{5e, e, e\} = e \end{aligned}$$

and then $v_p(\omega'') = v_p(s\zeta'' + t) \ge 3e$, which says that $P'' \in E_p^{(e)}(\mathbb{Q})$ as desired.

(2) First note that the computation in the proof of (1) above gives the congruence

$$z(P + P') = -\zeta'' \equiv z(P) + z(P') \bmod p^{5e}\,.$$

By induction, we obtain from this that $z(mP) \equiv mz(P) \bmod p^{5e}$. We now show the claim by induction on $k = v_p(m)$.

$k = 0$: Then $v_p(z(mP)) = v_p(mz(P)) = e = e + k$.

$k > 0$: We have that $z(pP) \equiv pz(P) \bmod p^{5e}$, so $v_p(z(pP)) = e + 1$. Using the induction hypothesis for pP, we obtain

$$v_p(z(mP)) = v_p(z(\tfrac{m}{p} \cdot pP)) = v_p(z(pP)) + v_p(m/p) = (e + 1) + (k - 1) = e + k\,.$$

(3) Let $P \in E_p^{(1)}(\mathbb{Q})_{\text{tors}}$ and assume that $P \neq O$. Then $e = v_p(z(P)) \in \mathbb{Z}_{\ge 1}$. Since P is a torsion point, there is $m \in \mathbb{Z}_{\ge 2}$ such that $mP = O$. Then (2) shows that $v_p(z(mP)) = e + v_p(m) < \infty$, so $mP \neq O$. This contradiction shows that we must have $P = O$. □

This proof works in the same way over the field $\mathbb{Q}_p$ of p-adic numbers in place of $\mathbb{Q}$. In this situation one can use the completeness of $\mathbb{Q}_p$ and the theory of *formal groups* (see [Si1, Ch. IV]) to give a very precise description of the group $E^{(1)}(\mathbb{Q}_p)$: If E is given by a short Weierstrass equation, then $E^{(1)}(\mathbb{Q}_p)$ is isomorphic to the additive group $p\mathbb{Z}_p$, and the isomorphism maps $E^{(e)}(\mathbb{Q}_p)$ to $p^e\mathbb{Z}_p$.

REM
Theorem 18.3 for long Weierstrass equations

18.4. **Remark.** If E is given by a long Weierstrass equation

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, \ldots, a_6 \in \mathbb{Z}$ instead of a short one as in Theorem 18.3, then Lemma 18.2 remains valid. One gets more terms in the equation, but this does not affect the estimates for the p-adic valuation of s and t. The proof of Theorem 18.3 needs to be modified as follows. The congruence for $z(P + P')$ holds in the weaker form

$$z(P + P') \equiv z(P) + z(P') \bmod p^{2e} \quad \text{resp.} \quad \bmod p^{3e}\,, \quad \text{when } a_1 = 0.$$

This shows that claim (1) remains valid in general. Claim (2) holds generally for $p \nmid m$, but for arbitrary m we need to assume in addition $e \ge 2$ or $a_1 = 0$. Since the proof only uses the p-adic valuation, one can weaken the integrality assumption on the coefficients to $v_p(a_j) \ge 0$. If $p \neq 2$, then we can complete the square and thus obtain an equation for an isomorphic curve E' that has $a_1 = a_3 = 0$ and $v_p(a_j) \ge 0$ for $j = 2, 4, 6$, such that the associated isomorphism $E \to E'$ maps $E_p^{(e)}(\mathbb{Q})$ to ${E'}_p^{(e)}(\mathbb{Q})$ (the x-coordinate and hence also its p-adic valuation is unchanged). Claims (2) and (3) therefore remain valid for $p \ge 3$ (since they hold for E'). In the case $p = 2$ we have the weaker statements $E_2^{(2)}(\mathbb{Q})_{\text{tors}} = \{O\}$ and $E_2^{(1)}(\mathbb{Q})_{\text{tors}} \subset E(\mathbb{Q})[2]$ in place of claim (3).

That this cannot be improved is shown by the example

$$E \colon y^2 + xy + y = x^3 + x^2 - 110x - 880\,;$$

this elliptic curve has the point $(\frac{51}{4}, -\frac{55}{8})$ of order 2, which lies in $E_2^{(1)}(\mathbb{Q})$. ♠

Claim (3) of the theorem gives the integrality of the torsion points. We add another statement that shows in particular that $E(\mathbb{Q})_{\text{tors}}$ is finite.

THM Nagell-Lutz theorem

18.5. **Theorem.** *Let $E \colon y^2 = x^3 + ax + b$ be an elliptic curve over $\mathbb{Q}$ with $a, b \in \mathbb{Z}$ and let $O \neq P = (\xi, \eta) \in E(\mathbb{Q})_{\text{tors}}$. Then $\xi, \eta \in \mathbb{Z}$, and either $\eta = 0$ or else η^2 divides $4a^3 + 27b^2$.*

Proof. If $\xi \notin \mathbb{Z}$ or $\eta \notin \mathbb{Z}$, then there would be a prime p and $e \geq 1$ such that $v_p(\xi) = -2e$, $v_p(\eta) = -3e$, hence

$$P \in E_p^{(e)}(\mathbb{Q}) \cap E(\mathbb{Q})_{\text{tors}} \subset E_p^{(1)}(\mathbb{Q})_{\text{tors}} = \{O\}$$

by Theorem 18.3 (3), contradicting the assumption $P \neq O$. Hence ξ and η are integers. We now assume that $\eta \neq 0$; we then have to show that $\eta^2 \mid 4a^3 + 27b^2$. Since $\eta \neq 0$, we have $2P \neq O$. $2P$ is again a torsion point, so it has integral coordinates by what we have just shown. We now prove the following more general statement.

If $P = (\xi, \eta)$ and $2P$ are both integral points on E, then η^2 divides $4a^3 + 27b^2$.

For the proof we recall that

$$x(2P) = \frac{\xi^4 - 2a\xi^2 - 8b\xi + a^2}{4(\xi^3 + a\xi + b)}\,.$$

$x(2P) \in \mathbb{Z}$ then implies that $\eta^2 = \xi^3 + a\xi + b$ divides $\xi^4 - 2a\xi^2 - 8b\xi + a^2$. Substituting ξ for x in the relation

$$(3x^2 + 4a)(x^4 - 2ax^2 - 8bx + a^2) - (3x^3 - 5ax - 27b)(x^3 + ax + b) = 4a^3 + 27b^2$$

gives the claim since η^2 divides both terms on the left. ❑

COR $E(\mathbb{Q})_{\text{tors}}$ is finite

18.6. **Corollary.** *Let E be an elliptic curve over $\mathbb{Q}$. Then the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ is finite.*

Proof. E is isomorphic to a curve that is given by a short Weierstrass equation. By scaling x and y we can achieve that the equation has integral coefficients. So without loss of generality, we can assume that E has this form. Then the Nagell-Lutz Theorem 18.5 shows that there are only finitely many possibilities for the y-coordinate of a torsion point $P \neq O$ (namely $y = 0$ and the $y \in \mathbb{Z}$ such that $y^2 \mid 4a^3 + 27b^2 \neq 0$; note that $\Delta(E) = -16(4a^3 + 27b^2)$). For each given y there are at most three possible values of x such that $(x, y) \in E$. ❑

The Nagell-Lutz Theorem gives an algorithm that allows to determine the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ in concrete cases. This requires factoring the discriminant of E, however. We first determine the points with $y = 0$ (these are exactly the points of order 2). From the factorization of $4a^3 + 27b^2$ we deduce the list of η such that $\eta^2 \mid 4a^3 + 27b^2$; for each such η we determine the integral roots ξ of $x^3 + ax + b - \eta^2$, and for each integral point $P = (\xi, \eta)$ found in this way, we compute $2P$, $4P$, $8P$, …, until we either obtain the point O or a point that has already appeared (in these cases, $P \in E(\mathbb{Q})_{\text{tors}}$), or we obtain a point that is not integral (then $P \notin E(\mathbb{Q})_{\text{tors}}$).

18.7. Example. We consider the elliptic curve

EX torsion subgroup

$$E\colon y^2 = x^3 - x + 1$$

with $a = -1$ and $b = 1$. Then $4a^3 + 27b^2 = 23$. The possible y-coordinates of nontrivial torsion points are therefore $y = -1, 0, 1$. This leads to the candidate points

$$(-1, \pm 1), \quad (0, \pm 1), \quad (1, \pm 1)\,.$$

We compute $2P$, $4P$, ... for each of these points (up to negation):

$$2 \cdot (-1,1) = (3,-5)\,, \qquad 2 \cdot (3,-5) = \left(\tfrac{19}{25}, -\tfrac{103}{125}\right) \notin E(\mathbb{Q})_{\text{tors}}$$
$$2 \cdot (0,1) = \left(\tfrac{1}{4}, -\tfrac{7}{8}\right) \notin E(\mathbb{Q})_{\text{tors}}$$
$$2 \cdot (1,1) = (-1,1) \notin E(\mathbb{Q})_{\text{tors}}$$

This shows that $E(\mathbb{Q})_{\text{tors}} = \{O\}$. ♣

18.8. Example. We consider the elliptic curve

EX torsion subgroup

$$E\colon y^2 = x^3 - 1386747x + 368636886\,.$$

We compute

$$4a^3 + 27b^2 = -69981157641830400000 = -2^{16} \cdot 3^{20} \cdot 5^4 \cdot 7^2\,.$$

The possible y-coordinates of nontrivial torsion points are therefore 0 and the divisors of $2^8 \cdot 3^{10} \cdot 5^2 \cdot 7$; there are $2 \cdot (8+1) \cdot (10+1) \cdot (2+1) \cdot (1+1) = 1188$ of these. For each of these choices we find all points in $E(\mathbb{Q})$ with that y-coordinate. This gives the following list.

$$(-1293, 0), \quad (282, 0), \quad (1011, 0),$$
$$(147, \pm 12960), \quad (1227, \pm 22680), \quad (-285, \pm 27216),$$
$$(-933, \pm 29160), \quad (2307, \pm 97200), \quad (8787, \pm 816480)\,.$$

The first three points have order 2. We double the remaining points repeatedly until we obtain either a torsion point or a non-integral point.

$$\begin{aligned}
2 \cdot (147, 12960) &= (2307, 97200) \\
2 \cdot (2307, 97200) &= (1011, 0) \in E(\mathbb{Q})_{\text{tors}} \\
2 \cdot (1227, 22680) &= (2307, -97200) \in E(\mathbb{Q})_{\text{tors}} \\
2 \cdot (-285, 27216) &= (1011, 0) \in E(\mathbb{Q})_{\text{tors}} \\
2 \cdot (-933, 29160) &= (2307, -97200) \in E(\mathbb{Q})_{\text{tors}} \\
2 \cdot (8787, 816480) &= (2307, 97200) \in E(\mathbb{Q})_{\text{tors}}
\end{aligned}$$

This shows that all points that we have found are indeed torsion points. We also see that $E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/8\mathbb{Z}$ with generators (e.g.) $(282, 0)$ and $(147, 12960)$. ♣

We will see in the next section how one can determine $E(\mathbb{Q})_{\text{tors}}$ without factoring the discriminant.

Later in this course we will discuss Mordell's Theorem (even though we will not prove it in all cases), which says that the group $E(\mathbb{Q})$ is *finitely generated*. This also implies that $E(\mathbb{Q})_{\text{tors}}$ is finite.

There is another theorem, due to Siegel,[6] which (as a special case) says that an elliptic curve over $\mathbb{Q}$ can have only finitely many integral points. Combined with

C.L. Siegel
1896–1981
Foto © MFO

[6] C.L. Siegel, *Über einige Anwendungen diophantischer Approximationen,* Abhandlungen Akad. Berlin 1929, No. 1, 70 S. (1929).

Theorem 18.3 (3) this shows again that $E(\mathbb{Q})_{\mathrm{tors}}$ is finite. Siegel's Theorem is quite deep, however (and uses Mordell's Theorem).

19. Good and bad reduction

If an elliptic curve

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with coefficients $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$ is given, then for a prime p, we can consider the curve

$$\bar{E}\colon y^2 + \bar{a}_1xy + \bar{a}_3y = x^3 + \bar{a}_2x^2 + \bar{a}_4x + \bar{a}_6\,,$$

where $\bar{a} \in \mathbb{F}_p$ denotes the residue class of a mod p. If p does not divide the discriminant $\Delta(E)$, then $\Delta(\bar{E}) = \overline{\Delta(E)} \neq 0$, and so $\bar{E}$ is an elliptic curve over $\mathbb{F}_p$.

Instead of E, we can consider a curve E' that is isomorphic to E over $\mathbb{Q}$ and whose equation also has integral coefficients. Then $\Delta(E') = u^{12}\Delta(E)$ for some $u \in \mathbb{Q}^\times$. In particular, it is possible that p divides $\Delta(E)$ but not $\Delta(E')$. This motivates the following definition.

DEF minimal Weierstrass equation

19.1. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$. A Weierstrass equation

$$E'\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

is a *minimal Weierstrass equation for E*, if E' is isomorphic to E over $\mathbb{Q}$, we have $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$, and $|\Delta(E')|$ is minimal among all such equations. Then $\Delta(E')$ is called the *minimal discriminant* of E. ♢

The minimality condition is equivalent to "$\Delta(E')$ divides the discriminant of every integral Weierstrass equation of an elliptic curve isomorphic to E". This minimality therefore also holds with respect to divisibility. The minimal discriminant is uniquely determined (exercise).

DEF good/bad reduction

19.2. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$; we assume that E is given by a minimal Weierstrass equation. Let p be a prime number. We say that E has *good reduction at p*, if p does not divide the (minimal) discriminant $\Delta(E)$; the elliptic curve $\bar{E}$ as above then is the *reduction of E mod p*. Otherwise, we say that E has *bad reduction at p*. ♢

We will now show that there is a relationship between the two groups $E(\mathbb{Q})$ and $\bar{E}(\mathbb{F}_p)$. We begin by showing that there is a reasonable map $\mathbb{P}^2(\mathbb{Q}) \to \mathbb{P}^2(\mathbb{F}_p)$.

LEMMA reduction of points in $\mathbb{P}^2(\mathbb{Q})$

19.3. **Lemma.** *Let $P = (\xi : \eta : \zeta) \in \mathbb{P}^2(\mathbb{Q})$. Then we can write P in the form $P = (\xi' : \eta' : \zeta')$ with $\xi', \eta', \zeta' \in \mathbb{Z}$ such that $\operatorname{ggT}(\xi', \eta', \zeta') = 1$. Let p be a prime number. Then the map*

$$\rho_p\colon \mathbb{P}^2(\mathbb{Q}) \longrightarrow \mathbb{P}^2(\mathbb{F}_p)\,, \quad P \longmapsto (\bar{\xi}' : \bar{\eta}' : \bar{\zeta}')$$

(where $\mathbb{Z} \to \mathbb{F}_p$, $a \mapsto \bar{a}$, is the canonical epimorphism) is well-defined.

If $G \subset \mathbb{P}^2_{\mathbb{Q}}$ is a line, then $\rho_p(G(\mathbb{Q})) = \bar{G}(\mathbb{F}_p)$ with a line $\bar{G} \subset \mathbb{P}^2_{\mathbb{F}_p}$.

Proof. We first scale the coordinates of P with a common denominator; we thus obtain an expression for P with integral coordinates. Then we scale by the inverse of the gcd of these coordinates; this results in an expression with coprime coordinates as claimed. If we scale these coordinates by $\lambda \in \mathbb{Z}$, then their gcd scales with $|\lambda|$. This implies that the coordinates are unique up to scaling by -1.

The point $(\bar{\xi}' : \bar{\eta}' : \bar{\zeta}') \in \mathbb{P}^2(\mathbb{F}_p)$ is then well-defined, since at least one of the coordinates is nonzero (this is because $p \nmid \operatorname{ggT}(\xi', \eta', \zeta')$), and the coordinates are unique up to scaling (by -1). Then also ρ_p is well-defined.

Now let $G\colon ax + by + cz = 0$ be a line in $\mathbb{P}^2_{\mathbb{Q}}$. In a similar way as we did for the coordinates of P, we can scale the equation of G in such a way that we have $a, b, c \in \mathbb{Z}$ with $\gcd(a, b, c) = 1$. Let then $\bar{G}\colon \bar{a}x + \bar{b}y + \bar{c}z = 0$ in $\mathbb{P}^2_{\mathbb{F}_p}$. Plugging in ξ', η', ζ' and reducing mod p shows that $\rho_p(G(\mathbb{Q})) \subset \bar{G}(\mathbb{F}_p)$. If conversely $(\bar{\xi} : \bar{\eta} : \bar{\zeta}) \in \bar{G}(\mathbb{F}_p)$, then let ξ', η', ζ' be arbitrary integers in the respective residue classes. Then $a\xi' + b\eta' + c\zeta' = pd$ with $d \in \mathbb{Z}$. Since $\gcd(a, b, c) = 1$, there are $r, s, t \in \mathbb{Z}$ such that $ra + sb + tc = 1$. Setting

$$(\xi, \eta, \zeta) := (\xi' - prd, \eta' - psd, \zeta' - ptd)\,,$$

we then have that $P := (\xi : \eta : \zeta) \in G(\mathbb{Q})$ and $\rho_p(P) = (\bar{\xi} : \bar{\eta} : \bar{\zeta})$. ❑

19.4. **Corollary.** *Let E be an elliptic curve over $\mathbb{Q}$ given by an equation with integral coefficients. Let p be a prime number such that $p \nmid \Delta(E)$ and let $\bar{E}$ be the elliptic curve obtained by reducing E mod p. Then the restriction of ρ_p to $E(\mathbb{Q})$ gives a group homomorphism*

COR
reduction
homomorphism

$$\rho_{p,E}\colon E(\mathbb{Q}) \longrightarrow \bar{E}(\mathbb{F}_p)\,.$$

We have that $\ker \rho_{p,E} = E_p^{(1)}(\mathbb{Q})$.

The last statement explains the name "kernel of reduction" for $E_p^{(1)}(\mathbb{Q})$.

Proof. Substituting suitably scaled integral coordinates into the equation of E and reduction mod p show that $\rho_p(E(\mathbb{Q})) \subset \bar{E}(\mathbb{F}_p)$; hence $\rho_{p,E}$ is at least well-defined as a map. We still have to show that $\rho_{p,E}$ is a group homomorphism. It is clear that $\rho_{p,E}(O) = O$. Let $P_1, P_2, P_3 \in E(\mathbb{Q})$ such that $P_1 + P_2 + P_3 = O$. Then P_1, P_2, P_3 are the three points of intersection of E with a line G (counted with multiplicity). Lemma 19.3 implies that $\rho_{p,E}(P_1), \rho_{p,E}(P_2), \rho_{p,E}(P_3)$ are the three points of intersection of $\bar{E}$ with the line $\bar{G}$. Hence $\rho_{p,E}(P_1) + \rho_{p,E}(P_2) + \rho_{p,E}(P_3) = O$ in $\bar{E}(\mathbb{F}_p)$. This shows that $\rho_{p,E}$ is a group homomorphism.

For the proof of the last statement we can choose the projective coordinates of $P \in E(\mathbb{Q})$ to be coprime integers. Then

$$\begin{aligned} P = (\xi : \eta : \zeta) \in \ker \rho_{p,E} &\iff \rho_{p,E}(P) = O = (0 : 1 : 0) \\ &\iff p \mid \xi\,,\ p \nmid \eta\,,\ p \mid \zeta \\ &\overset{(*)}{\iff} v_p(\eta/\zeta) < 0 \\ &\iff P \in E_p^{(1)}(\mathbb{Q})\,. \end{aligned}$$

The direction "⇒" in the equivalence $(*)$ is trivial. For the converse, note that $v_p(\eta/\zeta) < 0$ implies that there is $e \geq 1$ such that $v_p(\xi/\zeta) = -2e$ and $v_p(\eta/\zeta) = -3e$ (this follows from a comparison of the valuations of the various terms in the equation of E; compare the previous section). Since $\min\{v_p(\xi), v_p(\eta), v_p(\zeta)\} = 0$, we must then have $v_p(\xi) = e$, $v_p(\eta) = 0$ and $v_p(\zeta) = 3e$ ($e = \infty$ is here allowed; this means that $P = O$). ❑

We can use the reduction homomorphism together with what we have established in the previous section to obtain statements on $E(\mathbb{Q})_{\text{tors}}$.

19.5. Theorem. *Let*

THM reduction of torsion

$$E\colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

be an elliptic curve over $\mathbb{Q}$ *with* $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$*. If* p *is a prime number such that* $p \nmid \Delta(E)$ *and if* $a_1 = 0$ *or* $p \geq 3$*, then*

$$\rho_{p,E}\big|_{E(\mathbb{Q})_{\mathrm{tors}}} \colon E(\mathbb{Q})_{\mathrm{tors}} \to \bar{E}(\mathbb{F}_p)$$

is injective.

Proof. Under the stated assumptions, we have that

$$\ker \rho_{p,E}\big|_{E(\mathbb{Q})_{\mathrm{tors}}} = \ker \rho_{p,E} \cap E(\mathbb{Q})_{\mathrm{tors}} = E_p^{(1)}(\mathbb{Q}) \cap E(\mathbb{Q})_{\mathrm{tors}} \overset{(*)}{=} \{O\}\,;$$

this implies the claim. We have shown the equality $(*)$ for short Weierstrass equations in Theorem 18.3. For the general case, see Remark 18.4. ❑

We can therefore realize $E(\mathbb{Q})_{\mathrm{tors}}$ as a subgroup of $\bar{E}(\mathbb{F}_p)$ for every prime p ($p \geq 3$ if $a_1 \neq 0$) such that $p \nmid \Delta(E)$. There are infinitely many such primes. Since $\bar{E}(\mathbb{F}_p)$ is finite, this gives another proof of the fact that $E(\mathbb{Q})_{\mathrm{tors}}$ is finite.

We can also use Theorem 19.5 to obtain a good bound for $\#E(\mathbb{Q})_{\mathrm{tors}}$ with little effort, since by Lagrange's Theorem, $\#E(\mathbb{Q})_{\mathrm{tors}}$ must divide $\#\bar{E}(\mathbb{F}_p)$.

19.6. Example. Let $E\colon y^2 = x^3 - x + 1$ with $\Delta(E) = -2^4 \cdot 23$. We can therefore apply Theorem 19.5 for example with $p = 3$ and $p = 5$. We then see that

EX $E(\mathbb{Q})_{\mathrm{tors}}$ is trivial

$$\#E(\mathbb{Q})_{\mathrm{tors}} \mid \#\bar{E}(\mathbb{F}_3) = 7 \qquad \text{and} \qquad \#E(\mathbb{Q})_{\mathrm{tors}} \mid \#\bar{E}(\mathbb{F}_5) = 8\,,$$

hence $E(\mathbb{Q})_{\mathrm{tors}}$ must be trivial. Since $E(\mathbb{Q})$ contains affine points (e.g., $(1,1)$), it follows that $E(\mathbb{Q})$ is infinite. (In fact, we have $E(\mathbb{Q}) \cong \mathbb{Z}$; the group is generated by $(1,1)$.) ♣

19.7. Example. Let

EX bound is not tight

$$E\colon y^2 + xy + y = x^3 + x^2 - 70x - 279\,;$$

then $\Delta(E) = -2 \cdot 19^5$. We obtain the following table.

p	3	5	7	11	13	17	23	29	31	37
$\#\bar{E}(\mathbb{F}_p)$	5	10	5	10	15	15	25	35	40	40

This shows that $\#E(\mathbb{Q})_{\mathrm{tors}} \in \{1, 5\}$. One can write down the polynomial whose zeros are the x-coordinates of the points of order 5 on E:

$$\begin{aligned}5x^{12} &+ 25x^{11} - 4284x^{10} - 112875x^9 - 904395x^8 + 1848750x^7 + 97164150x^6 \\ &+ 1084824520x^5 + 7387397375x^4 + 28604803425x^3 + 39626137350x^2 \\ &- 77025287125x - 228943289601\end{aligned}$$

and check that it has no integral roots. So we have $E(\mathbb{Q})_{\mathrm{tors}} = \{O\}$.

There is a reason why the bound for $\#E(\mathbb{Q})_{\mathrm{tors}}$ is not tight in this example. Namely, E is isogenous to

$$E'\colon y^2 + xy + y = x^3 + x^2 + 1\,,$$

which has $\#E'(\mathbb{Q})_{\text{tors}} = 5$. Theorem 19.5 then implies that $5 \mid \#\bar{E}'(\mathbb{F}_p)$ for all $p \geq 3$, $p \neq 19$. By Theorem 13.3 (this is the easy direction, which we have proved), $\#\bar{E}(\mathbb{F}_p) = \#\bar{E}'(\mathbb{F}_p)$, since $\bar{E}$ and $\bar{E}'$ are also isogenous. This shows that

$$\operatorname{ggT}\bigl(\{\#\bar{E}(\mathbb{F}_p) \mid p \geq 3 \text{ prime}, p \neq 19\}\bigr) = 5\,,$$

and so the bound *cannot* be tight. ♣

B. Mazur
* 1937
Foto © MFO

The Hasse bound for the number of $\mathbb{F}_p$-rational points on an elliptic curve (Theorem 13.2) implies the general bound

$$\#E(\mathbb{Q})_{\text{tors}} \leq \min\bigl\{\lfloor(\sqrt{p}+1)^2\rfloor \bigm| p \geq 3 \text{ prime}, p \nmid \Delta(E)\bigr\}\,.$$

So when the torsion subgroup is "large", E must have bad reduction at all "small" primes. (For $2 \nmid \Delta(E)$ one obtains the bound $\#E(\mathbb{Q})_{\text{tors}} \leq 10$.)

On the other hand, Barry Mazur has shown in the 1970s that the torsion subgroup of $E(\mathbb{Q})$ cannot be arbitrarily large.[7]

THM
Mazur's Theorem

19.8. Theorem. *Let E be an elliptic curve over $\mathbb{Q}$. Then*

$$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/d\mathbb{Z} \qquad \text{with } 1 \leq d \leq 10 \text{ or } d = 12,$$

or

$$E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2d\mathbb{Z} \qquad \text{with } 1 \leq d \leq 4.$$

Each of these possible group structure occurs for infinitely many pairwise non-isomorphic elliptic curves over $\mathbb{Q}$.

Since ± 1 are the only roots of unity in $\mathbb{Q}$, the existence of the Weil pairing implies that $E(\mathbb{Q})_{\text{tors}}$ must have one of the two forms given above (for some $d \geq 1$).

It is not hard to show the second statement, since one can write down explicit families of such curves that depend on one free parameter.

To study the question whether a point $P \in E(\mathbb{Q})$ of order d can exist, one can use (for $d \geq 4$) the "normal form"

$$E\colon y^2 + u\,xy + v\,y = x^3 + v\,x^2$$

with $u, v \in \mathbb{Q}$; here $P = (0,0)$. The condition that P has order d results in an equation $P_d(u,v)$ with a polynomial P_d. This equation defines an affine plane curve whose rational points correspond either to pairs (E,P) consisting of an elliptic curve E over $\mathbb{Q}$ and a point $P \in E(\mathbb{Q})$ of order d or else lead to a curve that is not smooth (and therefore is not an elliptic curve). There exists a smooth projective (not necessarily plane) curve $X_1(d)$ that is defined over $\mathbb{Q}$ and is birational over $\mathbb{Q}$ to this affine curve. The main statement in Mazur's Theorem above then follows from the fact that the rational points on $X_1(d)$ for d a prime ≥ 11 and for $d = 14$, 15, 16, 18, 20, 21, 24, 25, 27, 35, 49 all correspond to singular curves E. (Mazur's contribution was to show this for all prime $d \geq 11$.)

For example, $X_1(11)$ is itself an elliptic curve, and one can show that it has exactly five rational points, none of which correspond to an elliptic curve with a point of order 11. For large d, the genus of $X_1(d)$ grows quickly, and the proof in the general case requires very deep methods.

[7] B. Mazur, *Modular curves and the Eisenstein ideal,* Inst. Hautes Études Sci. Publ. Math. **47**, 33–186 (1978).

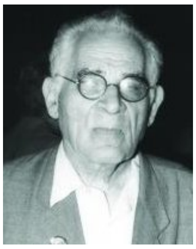
20. Mordell's Theorem

Our next goal is to prove the following theorem (at least in a special case).

20.1. **Theorem.** *Let E be an elliptic curve over $\mathbb{Q}$. Then $E(\mathbb{Q})$ is a finitely generated abelian group.*

THM
Mordell's
Theorem

This theorem was first proved by Mordell.[8] The statement was generalized by Weil a few years later in his PhD thesis to Jacobian varieties of algebraic curves over arbitrary algebraic number fields (i.e., finite field extensions of $\mathbb{Q}$).[9] (Elliptic curves are their own Jacobian varieties.) Therefore this theorem is usually referred to as the *Mordell-Weil Theorem*. (Mordell himself seemed to be quite unhappy about that.)



L.J. Mordell
1888 – 1972
Foto © MFO

We will, so to speak, put the cart before the horse and reduce the proof of the theorem to several other claims.

20.2. **Theorem.** *Let G be an (additively written) abelian group, $m \in \mathbb{Z}_{\geq 2}$ and $h\colon G \to \mathbb{R}_{\geq 0}$ a map with the following properties.*

THM
abelian group
is finitely
generated

(1) *G/mG is finite.*

(2) *For every $B > 0$, the set $\{g \in G \mid h(g) \leq B\}$ is finite.*

(3) *There is some $C > 0$ such that $h(mg) \geq m^2 h(g) - C$ for all $g \in G$.*

(4) *For each $g \in G$ there is some $c_g > 0$ such that $h(g + g') \leq 2h(g') + c_g$ for all $g' \in G$.*

Then G is finitely generated.

20.3. **Definition.** A map $h\colon G \to \mathbb{R}_{\geq 0}$ with the properties (2), (3) (for some $m \geq 2$) and (4) is a *height function* on G. ♢

DEF
height
function

Proof. Let $g_i \in G$ for $i = 1, \dots, k$ be representatives of the by (1) finitely many cosets in G/mG, and let

$$\gamma = \frac{C + \max\{c_{-g_i} \mid i = 1, \dots, k\}}{m^2 - 2} > 0\,.$$

We show that G is generated by the (according to (2)) finite set

$$M = \{g_i \mid i = 1, \dots, k\} \cup \{g \in G \mid h(g) \leq \gamma\}\,.$$

We argue by contradiction and assume that this is not the case. Then there exist elements $g \in G$ such that $g \notin \langle M \rangle$. Since by (2) there are only finitely many elements $g' \in G$ with $h(g') \leq h(g)$, we can assume that g has minimal height $h(g)$ among all such counterexamples. We then certainly have $g \notin M$, which implies $h(g) > \gamma$. Let now g_i be the chosen representative of the coset $g + mG \in G/mG$. Then there is $g' \in G$ such that $g = g_i + mg'$. Properties (3) and (4) then imply that

$$2h(g) + c_{-g_i} \geq h(g - g_i) = h(mg') \geq m^2 h(g') - C$$

[8] L.J. Mordell, *On the rational solutions of the indeterminate equations of the third and fourth degrees,* Cambr. Phil. Soc. Proc. **21**, 179–192 (1922).

[9] A. Weil, *L'arithmétique sur les courbes algébriques,* Acta Math. **52**, 281–315 (1929).

and therefore

$$h(g') \le h(g) - \frac{(m^2-2)h(g) - (C + c_{-g_i})}{m^2} \le h(g) - \frac{m^2-2}{m^2}(h(g) - \gamma) < h(g)\,.$$

Since g was a counterexample of minimal height, we must have $g' \in \langle M \rangle$. But then we also have $g = g_i + mg' \in \langle M \rangle$ (since $g_i \in M$). This is a contradiction, which shows that our assumption must be false. So we see that $G = \langle M \rangle$ as desired. ❑

Statement (1) above is known in the context of elliptic curves (or abelian varieties), i.e., with $G = E(\mathbb{Q})$, as the *weak Mordell-Weil Theorem.* It is clear that (1) is a necessary condition for the finite generation of G. This condition is not sufficient, however, as can be seen in examples like the additive groups of $\mathbb{Q}$ or $\mathbb{R}$. We will look at this weak theorem of Mordell (with $m = 2$) later. In this section we will show that there exists a height function on $E(\mathbb{Q})$.

20.4. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$. We set

DEF height function on $E(\mathbb{Q})$

$$h\colon E(\mathbb{Q}) \longrightarrow \mathbb{R}_{\ge 0}\,, \quad P \longmapsto \begin{cases} 0\,, & P = O, \\ \log\max\{|u|, |v|\}\,, & x(P) = u/v \text{ with } u \perp v. \end{cases} \quad \diamondsuit$$

Here log is the natural logarithm. The height $h(P)$ can be seen as a measure for the amount of space that one needs to write down (the x-coordinate of) P. If we set $x(O) = 1/0$, then the case distinction in the definition is not necessary. This is compatible with the interpretation of the x-coordinate map as the morphism $x\colon E \to \mathbb{P}^1$, $(\xi : \eta : \zeta) \mapsto (\xi : \zeta)$.

We will now show that h is indeed a height function on $E(\mathbb{Q})$.

20.5. **Theorem.** *Let $E\colon y^2 = x^3 + ax + b$ be an elliptic curve over $\mathbb{Q}$ with $a, b \in \mathbb{Z}$. Then h as in Definition 20.4 is a height function on $E(\mathbb{Q})$ (for $m = 2$).*

THM height function on $E(\mathbb{Q})$

Proof. We verify the three relevant properties in Theorem 20.2.

(2) If $P \in E(\mathbb{Q})$ such that $h(P) \le B$, then either $P = O$, or else $x(P) = u/v$ with $u, v \in \mathbb{Z}$, $|u|, |v| \le e^B$. There are therefore only finitely many possibilities for the x-coordinate of P, and there are at most two possible y-coordinates for each x-coordinate. This shows that there are only finitely many such points.

(3) We have to show that there is $C > 0$ such that

$$h(2P) \ge 4h(P) - C \qquad \text{for all } P \in E(\mathbb{Q}).$$

We use the duplication formula

$$x(2P) = \frac{x(P)^4 - 2ax(P)^2 - 8bx(P) + a^2}{4(x(P)^3 + ax(P) + b)}\,.$$

Writing $x(P) = u/v$ with $u \perp v$, we then have

$$x(2P) = \frac{u^4 - 2au^2v^2 - 8buv^3 + a^2v^4}{4(u^3 + auv^2 + bv^3)v} =: \frac{F_1(u,v)}{F_2(u,v)}\,, \tag{20.1}$$

where $F_1(u,v), F_2(u,v) \in \mathbb{Z}$, hence

$$h(2P) = \log\max\{|F_1(u,v)|, |F_2(u,v)|\} - \log g\,,$$

where g is the gcd of $F_1(u,v)$ and $F_2(u,v)$. Now we observe that

$$G_1(u,v)\cdot F_1(u,v)+G_2(u,v)\cdot F_2(u,v)=4(4a^3+27b^2)u^7 \quad \text{and}$$
$$H_1(u,v)\cdot F_1(u,v)+H_2(u,v)\cdot F_2(u,v)=4(4a^3+27b^2)v^7$$

with

$$\begin{aligned}
G_1(u,v)&=(16a^3+108b^2)u^3-4a^2bu^2v+(12a^4+88ab^2)uv^2+(12a^3b+96b^3)v^3\,,\\
G_2(u,v)&=a^2bu^3+(5a^4+32ab^2)u^2v+(26a^3b+192b^3)uv^2-(3a^5+24a^2b^2)v^3\,,\\
H_1(u,v)&=(12u^2+16av^2)v\,,\\
H_2(u,v)&=-(3u^3-5auv^2-27bv^3)\,.
\end{aligned}$$

Since $u\perp v$, this implies that $g \mid 4(4a^3+27b^2)=-\Delta(E)/4\neq 0$. The triangle inequality implies that there is a constant A such that

$$\max\{|G_1(u,v)|,|G_2(u,v)|,|H_1(u,v)|,|H_2(u,v)|\}\le A\max\{|u|,|v|\}^3\,.$$

We conclude that

$$\max\{|u|,|v|\}^7\le\frac{2A}{4|4a^3+27b^2|}\max\{|u|,|v|\}^3\max\{|F_1(u,v)|,|F_2(u,v)|\}\,.$$

From this and $|g|\le 4|4a^3+27b^2|$ we finally obtain that $h(2P)\ge 4h(P)-C$ with $C=\log(2A)$.

(4) We have to show that for every $P\in E(\mathbb{Q})$ there is a constant $c_P>0$ such that $h(P+Q)\le 2h(Q)+c_P$ for all $Q\in E(\mathbb{Q})$. This is clear when $P=O$, and the statement holds for $Q=O$ whenever $c_P\ge h(P)$. It also holds for $Q=-P$, and it holds for $Q=P$ whenever $c_P\ge h(2P)-2h(P)$. We can therefore assume that $P\neq O$ and $Q\neq O,P,-P$. We write $P=(\xi_P:\eta_P:\zeta_P)$ and $Q=(\xi_Q:\eta_Q:\zeta_Q)$ with $\xi_P,\zeta_P,\xi_Q,\zeta_Q\in\mathbb{Z}$ and $\xi_P\perp\zeta_P$, $\xi_Q\perp\zeta_Q$. Note that η_P and η_Q are *not* integral in general. However, multiplying the relation

$$\eta_Q^2\zeta_Q=\xi_Q^3+a\xi_Q\zeta_Q^2+b\zeta_Q^3$$

by ζ_Q, we see that $\eta_Q\zeta_Q\in\mathbb{Z}$ and

$$|\eta_Q\zeta_Q|\le\sqrt{1+|a|+|b|}\max\{|\xi_Q|,|\zeta_Q|\}^2 \tag{20.2}$$

(and similarly for P). We have

$$\begin{aligned}
x(P+Q)&=\Bigl(\frac{y(P)-y(Q)}{x(P)-x(Q)}\Bigr)^2-x(P)-x(Q)\\
&=\frac{(\xi_P\xi_Q+a\zeta_P\zeta_Q)(\xi_P\zeta_Q+\zeta_P\xi_Q)+2b\zeta_P^2\zeta_Q^2-2\eta_P\zeta_P\eta_Q\zeta_Q}{(\xi_P\zeta_Q-\zeta_P\xi_Q)^2}=:\frac{N}{D}\,.
\end{aligned}$$

The numerator N and denominator D of this fraction are integers. Equation (20.2) together with the triangle inequality yields the estimate

$$|N|,|D|\le 4(1+|a|+|b|)\max\{|\xi_P|,|\zeta_P|\}^2\max\{|\xi_Q|,|\zeta_Q|\}^2\,.$$

This implies that

$$\begin{aligned}
h(P+Q)&\le\log\max\{|N|,|D|\}\\
&\le\log\bigl(4(1+|a|+|b|)\bigr)+2\log\max\{|\xi_P|,|\zeta_P|\}+2\log\max\{|\xi_Q|,|\zeta_Q|\}\\
&=\log\bigl(4(1+|a|+|b|)\bigr)+2h(P)+2h(Q)\,;
\end{aligned}$$

the desired statement therefore holds with

$$c_P=\max\bigl\{\log\bigl(4(1+|a|+|b|)\bigr)+2h(P),h(2P)-2h(P)\bigr\}\,.$$

We remark that one can extract from (20.1) in a similar way the estimate

$$h(2P) \le 4h(P) + \log \max\{(1 + |a|)^2 + 8|b|, 4(1 + |a| + |b|)\}\,.$$

This gives the more explicit value

$$c_P = 2h(P) + \log \max\{(1 + |a|)^2 + 8|b|, 4(1 + |a| + |b|)\}\,. \qquad ❑$$

Given this, the proof of Mordell's Theorem 20.1 is reduced to proving the weak Theorem of Mordell for $m = 2$, i.e., the statement that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite. We will look at this in the following sections.

20.6. **Remark.** The existence of a height function on $E(\mathbb{Q})$ gives another proof of the finiteness of $E(\mathbb{Q})_{\text{tors}}$, as follows. If $P \in E(\mathbb{Q})_{\text{tors}}$, then the set $\{2^n P \mid n \ge 0\}$ is finite; let $H = \max\{h(2^n P) \mid n \ge 0\}$ and choose n in such a way that $h(2^n P) = H$. Property (3) then gives that

REM $\#E(\mathbb{Q})_{\text{tors}}$ is finite

$$H \ge h(2^{n+1}P) \ge 4h(2^n P) - C = 4H - C \quad \Longrightarrow \quad H \le \frac{C}{3}$$

and therefore that $h(P) \le H \le C/3$. We conclude that

$$E(\mathbb{Q})_{\text{tors}} \subset \{P \in E(\mathbb{Q}) \mid h(P) \le C/3\}\,,$$

and the set on the right is finite by property (2).

Since one can enumerate the elements of the set on the right explicitly at least in principle (consider all $\xi = u/v$ such that $|u|, |v| \le e^{C/3}$ and check which of these are x-coordinates of points in $E(\mathbb{Q})$), this provides another (albeit not very efficient) algorithm for the determination of $E(\mathbb{Q})_{\text{tors}}$. Note that $e^{C/3} = \sqrt[3]{2A}$ with A as in the proof of Theorem 20.5; A can be expressed as an explicit polynomial in $|a|$ and $|b|$. ♠

Using similar arguments as in the proof of Theorem 20.5 one can show the following stronger statement.

20.7. **Theorem.** *Let E be an elliptic curve over $\mathbb{Q}$. Then there is a constant c_E that depends only on (the equation of) E such that for all points $P, Q \in E(\mathbb{Q})$ we have*

THM approximate parallelogram law

$$|h(P + Q) + h(P - Q) - 2h(P) - 2h(Q)| \le c_E\,.$$

In particular,

$$|h(2P) - 4h(P)| \le c_E\,.$$

This can be used for the following construction.

20.8. **Theorem.** *Let E be an elliptic curve over $\mathbb{Q}$ and let $P \in E(\mathbb{Q})$. Then there exists*

THM canonical height

$$\hat{h}(P) := \hat{h}_E(P) := \lim_{n\to\infty} \frac{h(2^n P)}{4^n} \in \mathbb{R}_{\ge 0}\,.$$

The function $\hat{h} \colon E(\mathbb{Q}) \to \mathbb{R}_{\ge 0}$ has the following properties.

(1) *There is γ_E such that $|\hat{h}(P) - h(P)| \le \gamma_E$ for all $P \in E(\mathbb{Q})$.*

(2) *For each $B \ge 0$, the set $\{P \in E(\mathbb{Q}) \mid \hat{h}(P) \le B\}$ is finite.*

(3) *For all $P, Q \in E(\mathbb{Q})$,*

$$\hat{h}(P + Q) + \hat{h}(P - Q) = 2\hat{h}(P) + 2\hat{h}(Q)\,.$$

(4) *For all $P \in E(\mathbb{Q})$ and all $m \in \mathbb{Z}$ we have $\hat{h}(mP) = m^2\hat{h}(P)$.*

(5) *For all $P \in E(\mathbb{Q})$ we have that $\hat{h}(P) = 0 \iff P \in E(\mathbb{Q})_{\mathrm{tors}}$.*

(6) *If $\phi\colon E \to E'$ is an isomorphism, then $\hat{h}_E = \hat{h}_{E'} \circ \phi$.*

(7) *$\hat{h}$ induces a positive definite quadratic form on $E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R}$.*

20.9. **Definition.** The function $\hat{h}_E$ above is the *canonical height* on E. ◊

DEF canonical height

"Canonical" since it does not depend on the given equation of E (and since it also has nice algebraic properties).

The canonical height is also a height function on $E(\mathbb{Q})$, with even better properties than an arbitrary height function.

Properties (4) and (1) then also imply that h is a height function on $E(\mathbb{Q})$ for *every* $m \geq 2$.

Proof. Theorem 20.7 implies that

$$\left|\frac{h(2^{n+1}P)}{4^{n+1}} - \frac{h(2^nP)}{4^n}\right| = \frac{1}{4^{n+1}}|h(2 \cdot 2^nP) - 4h(2^nP)| \leq \frac{c_E}{4^{n+1}}\,.$$

This in turn implies for $n \geq m$ that

$$\left|\frac{h(2^nP)}{4^n} - \frac{h(2^mP)}{4^m}\right| \leq \sum_{k=m}^{n-1}\left|\frac{h(2^{k+1}P)}{4^{k+1}} - \frac{h(2^kP)}{4^k}\right| \leq c_E \sum_{k=m}^{\infty} \frac{1}{4^{k+1}} = \frac{c_E}{3 \cdot 4^m}\,,$$

so $\bigl(4^{-n}h(2^nP)\bigr)_{n\geq 0}$ is a Cauchy sequence, and the limit exists.

(1) The consideration above shows in particular that $|4^{-n}h(2^nP) - h(P)| \leq c_E/3$. Taking the limit as $n \to \infty$ gives the claim (with $\gamma_E = c_E/3$).

(2) From $\hat{h}(P) \leq B$ we obtain with (1) that $h(P) \leq B + \gamma_E$. The claim therefore follows from the corresponding property of h.

(3) By Theorem 20.7, we have

$$\left|\frac{h(2^n(P+Q))}{4^n} + \frac{h(2^n(P-Q))}{4^n} - 2\frac{h(2^nP)}{4^n} - 2\frac{h(2^nQ)}{4^n}\right| \leq \frac{c_E}{4^n}\,.$$

The claim follows by letting n tend to infinity.

(4) This follows from (3) by induction.

(5) If $P \in E(\mathbb{Q})_{\mathrm{tors}}$, then $\{2^nP \mid n \geq 0\}$ is finite, so $h(2^nP)$ is bounded. The definition of $\hat{h}$ then implies that $\hat{h}(P) = 0$. Conversely, if $\hat{h}(P) = 0$, then by (4) $\hat{h}(mP) = 0$ for all $m \in \mathbb{Z}$. By (2) $\{Q \in E(\mathbb{Q}) \mid \hat{h}(Q) = 0\}$ is finite, so P has only finitely many distinct multiples. This implies that $P \in E(\mathbb{Q})_{\mathrm{tors}}$.

(6) Similarly as in the proof of Theorem 20.5, one sees that $h_{E'}(\phi(P)) \leq h_E(P) + c_\phi$ for some c_ϕ. In the same way, $h_E(\phi^{-1}(Q)) \leq h_{E'}(Q) + c_{\phi^{-1}}$. The difference between h_E and $h_{E'} \circ \phi$ is therefore bounded. As above, the claim follows by taking a suitable limit.

(7) First of all, (3) and (5) together imply that $\hat{h}$ induces a quadratic form on (the free abelian group) $E(\mathbb{Q})/E(\mathbb{Q})_{\mathrm{tors}}$. This gives us a quadratic form on the vector space $V_{\mathbb{Q}} := E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{Q}$ that takes strictly positive values away from $\mathbf{0}$. This implies that $\hat{h}$ is positive semi-definite on $V := E(\mathbb{Q}) \otimes_{\mathbb{Z}} \mathbb{R} = V_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{R}$. (2) implies that there is $\varepsilon > 0$ such that the torsion points are the only points $P \in E(\mathbb{Q})$ such that $\hat{h}(P) < \varepsilon$.

Let $\mathbf{0} \neq \boldsymbol{x} \in V$, then $\boldsymbol{x}$ is in $V' := \langle P_1, \ldots, P_n \rangle_{\mathbb{R}}$ for finitely many points $P_1, \ldots, P_n \in E(\mathbb{Q})$. The image of $\langle P_1, \ldots, P_n \rangle_{\mathbb{Z}} \subset E(\mathbb{Q})$ in V' is a lattice Λ with $\Lambda \otimes_{\mathbb{Z}} \mathbb{R} = V'$; we can assume that $\dim V' = n$. If $\hat{h}(\boldsymbol{x}) = 0$, then $\hat{h}$ would be given on V' with respect to a suitable basis by $x_1^2 + \ldots + x_m^2$ with $m < n$. In particular, the set $\{\boldsymbol{y} \in V' \mid \hat{h}(\boldsymbol{y}) < \varepsilon\}$ would be a centrally symmetric convex open set with infinite volume, which intersects Λ only at the origin. This contradicts Minkowski's Convex Body Theorem (see also the (German) lecture notes "'Diophantische Gleichungen"'). This shows that $\hat{h}(\boldsymbol{x}) > 0$, and so $\hat{h}$ is positive definite on V. ❑

21. The weak Mordell Theorem

In the following, we want to prove the weak Mordell Theorem in the form

$$E(\mathbb{Q})/2E(\mathbb{Q}) \textit{ is finite}$$

under the assumption that $E(\mathbb{Q})[2] \neq \{O\}$. Then there is a rational point T of order 2 on E. We can shift the x-coordinate of the given (short) Weierstrass equation so that $T = (0,0)$. Then the equation of E is

$$E\colon y^2 = x(x^2 + ax + b)\,. \tag{21.1}$$

Since E is an elliptic curve, we have $b \neq 0$ and $a^2 - 4b \neq 0$. There is an isogeny $\phi\colon E \to E'$ of degree 2, where

$$E'\colon y^2 = x(x^2 - 2ax + (a^2 - 4b))$$

and

$$\phi\colon E \longrightarrow E', \quad (x,y) \longmapsto \left(\frac{y^2}{x^2}, \frac{b - x^2}{x^2}\,y\right) = \left(\frac{x^2 + ax + b}{x}, \frac{b - x^2}{x^2}\,y\right) \quad \text{and}$$

$$\hat{\phi}\colon E' \longrightarrow E, \quad (x,y) \longmapsto \left(\frac{y^2}{4x^2}, \frac{a^2 - 4b - x^2}{8x^2}\,y\right).$$

We have $\ker(\phi) = \{O, T\}$ and $\ker(\hat{\phi}) = \{O', T'\}$, where $O' = (0 : 1 : 0) \in E'(\mathbb{Q})$ and $T' = (0,0) \in E'(\mathbb{Q})$; compare Example 11.13.

21.1. Lemma. *Let E, E', ϕ and $\hat{\phi}$ be as above. If the groups $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ and $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ are both finite, then $E(\mathbb{Q})/2E(\mathbb{Q})$ is also finite.*

LEMMA
$E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$
finite suffices

Proof. Since $2E(\mathbb{Q}) = (\hat{\phi} \circ \phi)(E(\mathbb{Q})) \subset \hat{\phi}(E'(\mathbb{Q}))$, we obtain a canonical epimorphism

$$\alpha\colon \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \longrightarrow \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))}\,.$$

Clearly, $\ker(\alpha) = \hat{\phi}(E'(\mathbb{Q}))/2E(\mathbb{Q})$. The homomorphism

$$\beta\colon \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} \longrightarrow \frac{E(\mathbb{Q})}{2E(\mathbb{Q})}\,, \qquad P' + \phi(E(\mathbb{Q})) \longmapsto \hat{\phi}(P') + 2E(\mathbb{Q})$$

is well-defined (since $\hat{\phi}(\phi(E(\mathbb{Q})) \subset 2E(\mathbb{Q})$), and $\operatorname{im}(\beta) = \ker(\alpha)$. We conclude that

$$\#\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} = \#\ker(\alpha)\cdot\#\operatorname{im}(\alpha) = \#\operatorname{im}(\beta)\cdot\#\frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} \le \#\frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))}\cdot\#\frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))}\,,$$

which implies the claim. ❑

To prove the weak Mordell Theorem, it therefore suffices to show that in the situation above, $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ is finite; the finiteness of the other group $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ follows in the same way (up to scaling x and y, we have $E = E''$).

Below we use the notation

$$\mathbb{Q}^{\times 2} := \{\alpha^2 \mid \alpha \in \mathbb{Q}^\times\}\,.$$

LEMMA $x(P)$ mod squares

21.2. **Lemma.** *The map*

$$\delta\colon E(\mathbb{Q}) \longrightarrow \frac{\mathbb{Q}^\times}{\mathbb{Q}^{\times 2}}\,, \qquad P \longmapsto \begin{cases} \mathbb{Q}^{\times 2}\,, & P = O\,, \\ b \cdot \mathbb{Q}^{\times 2}\,, & P = T\,, \\ x(P) \cdot \mathbb{Q}^{\times 2}\,, & \textit{else} \end{cases}$$

is a group homomorphism with kernel $\ker(\delta) = \hat{\phi}(E'(\mathbb{Q}))$.

Proof. We need to show that $\delta(P+Q) = \delta(P)\cdot\delta(Q)$ for $P, Q \in E(\mathbb{Q})$. This is clear when $P = O$ or $Q = O$ or $P = Q = T$ or $P + Q = O$ (using that $\delta(-P) = \delta(P)$). If (e.g.) $Q = T$ and $P \neq T$, then the claim follows from $x(P + T) = b/x(P)$:

$$\delta(P + T) = \frac{b}{x(P)} \cdot \mathbb{Q}^{\times 2} = x(P) \cdot b \cdot \mathbb{Q}^{\times 2} = \delta(P) \cdot \delta(T)\,.$$

If $P, Q \notin \{O, T\}$, but $P + Q = T$, then $Q = -P + T$, and the claim follows in the same way. We can therefore assume that P, Q and $P + Q$ are all distinct from O and T. Then $P, Q, R := -(P + Q)$ are the intersection points of E with a line $y = \lambda x + \mu$. This implies

$$x(x^2 + ax + b) - (\lambda x + \mu)^2 = (x - x(P))(x - x(Q))(x - x(R))\,.$$

Evaluating this at $x = 0$ shows that $x(P)x(Q)x(R) = \mu^2 \in \mathbb{Q}^{\times 2}$ (all three x-coordinates are nonzero). This is equivalent to

$$\delta(P + Q) = \delta(R) = \delta(P) \cdot \delta(Q)\,.$$

We now show that $\hat{\phi}(E'(\mathbb{Q})) \subset \ker(\delta)$. Let $P' \in E'(\mathbb{Q})$. If $P' \in \ker(\hat{\phi}) = \{O', T'\}$, then $\delta(\hat{\phi}(P')) = \delta(O) = \mathbb{Q}^{\times 2}$. If $x(\hat{\phi}(P')) = 0$, then we must have $y(P') = 0$ and $P' \neq T'$, so $x(P')$ is a rational root of $x^2 - 2ax + a^2 - 4b$. This shows that the discriminant $(2a)^2 - 4(a^2 - 4b) = 16b$ is a square, so b is a square, and we have $\delta(\hat{\phi}(P')) = \delta(T) = \mathbb{Q}^{\times 2}$. We now assume that $x(\hat{\phi}(P')) \notin \{0, \infty\}$. The explicit formula for $\hat{\phi}$ above shows that $x(\hat{\phi}(P'))$ is a square $\neq 0$, hence $\delta(\hat{\phi}(P')) = \mathbb{Q}^{\times 2}$.

It remains to show the reverse inclusion. So let $P \in \ker(\delta)$. If $P = O$, then $P \in \hat{\phi}(E'(\mathbb{Q}))$. If $P = T$, then b is a square, which implies (see above) that there is a point $P' = (\xi, 0) \neq T'$ in $E'(\mathbb{Q})$; then $\hat{\phi}(P') = T$. If $P \notin \{O, T\}$, then $x(P) = \alpha^2$ is a nonzero square. We then see from the equation of E that $\alpha^4 + a\alpha^2 + b$ is a square (namely, $(y(P)/\alpha)^2$). We are looking for a point $P' \in E'(\mathbb{Q})$ such that $\hat{\phi}(P') = P$. We substitute $y = 2\alpha x$ in the equation of E'; after canceling x, we obtain

$$x^2 - 2(a + 2\alpha^2)x + a^2 - 4b = 0\,.$$

We have

$$(a + 2\alpha^2)^2 - (a^2 - 4b) = 4(a\alpha^2 + \alpha^4 + b) = \Bigl(\frac{2y(P)}{\alpha}\Bigr)^2\,,$$

so the equation has a solution $\xi \in \mathbb{Q}$, and $P' = (\xi, 2\alpha\xi)$ is a rational point on E'. Then $\hat{\phi}(P') \in E(\mathbb{Q})$ is a point with the same x-coordinate as P, hence we have either $P = \hat{\phi}(P')$ or else $P = -\hat{\phi}(P') = \hat{\phi}(-P')$. ❑

21.3. Corollary. *Let E, E', ϕ, $\hat{\phi}$ be as above and let δ be as in Lemma 21.2. If δ has finite image, then $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ is also finite.*

COR
$\operatorname{im}(\delta)$ finite suffices

Proof. Lemma 21.2 implies that δ induces an isomorphism

$$\frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} = \frac{E(\mathbb{Q})}{\ker(\delta)} \xrightarrow{\cong} \operatorname{im}(\delta)\,. \qquad ❑$$

The idea for finishing the proof will now be to find a finite upper bound for $\operatorname{im}(\delta)$. i.e., a finite subgroup $S_{\hat{\phi}} \subset \mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ that contains $\operatorname{im}(\delta)$.

The group $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ itself is infinite. It is an $\mathbb{F}_2$-vector space of countably infinite dimension, with basis $(-1)\cdot\mathbb{Q}^{\times 2}$ and $p\cdot\mathbb{Q}^{\times 2}$ for each prime number p (this follows easily from the unique prime factorization in $\mathbb{Z}$). In particular, every coset contains a unique representative that is a squarefree integer.

We will now show that the squarefree integer that represents $\delta(P)$ can have only certain prime divisors.

Consider $P \in E(\mathbb{Q})$ with $P \notin \{O, T\}$. From the considerations at the beginning of Section 18 we know that there are integers r, s, t such that $r \perp t$, $s \perp t$ and $P = (r/t^2, s/t^3)$. Substituting this into the equation (21.1) of E and clearing denominators, we obtain

$$s^2 = r(r^2 + art^2 + bt^4)\,;$$

we also have that $\delta(P) = r \cdot \mathbb{Q}^{\times 2}$. We assume (as we may do) that a and b are integers.

Let p be a prime divisor of r such that $v_p(r)$ is odd. Since $v_p(s^2) = 2v_p(s)$ is even, p must divide $r^2 + art^2 + bt^4$. We therefore have that

$$p \mid \operatorname{ggT}(r, r^2 + art^2 + bt^4) = \operatorname{ggT}(r, bt^4) = \operatorname{ggT}(r, b) \mid b$$

(this uses $r \perp t$). We see that $\delta(P)$ lies in the subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ that is generated by the classes of -1 and the prime divisors of b. This remains true when $P = O$ (obviously) or $P = T$ (since $\delta(T) = b \cdot \mathbb{Q}^{\times 2}$).

21.4. Theorem. *Let E be an elliptic curve over $\mathbb{Q}$ such that $E(\mathbb{Q})[2] \neq \{O\}$. Then $E(\mathbb{Q})$ is finitely generated.*

THM
Mordell's Theorem

Proof. E is isomorphic to an elliptic curve of the form (21.1) with $a, b \in \mathbb{Z}$; we can therefore assume that E has this form. Then $b \neq 0$ and $a^2 - 4b \neq 0$. Therefore the subgroups

$$H = \langle (-1)\cdot\mathbb{Q}^{\times 2}, p\cdot\mathbb{Q}^{\times 2} \mid p \text{ prime}, p \mid b\rangle \quad \text{and}$$
$$H' = \langle (-1)\cdot\mathbb{Q}^{\times 2}, p\cdot\mathbb{Q}^{\times 2} \mid p \text{ prime}, p \mid a^2 - 4b\rangle$$

of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ are finite. In the discussion preceding the statement of the theorem, we have seen that $\operatorname{im}(\delta) \subset H$; in the same way, it follows that $\operatorname{im}(\delta') \subset H'$, where $\delta'\colon E'(\mathbb{Q}) \to \mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ is the analogous map. Corollary 21.3 then shows that $E(\mathbb{Q})/\hat{\phi}(E'(\mathbb{Q}))$ and $E'(\mathbb{Q})/\phi(E(\mathbb{Q}))$ are both finite. Then Lemma 21.1 implies that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite, which is the weak Mordell Theorem for E and $m = 2$. By Theorem 20.5, there exists a height function on $E(\mathbb{Q})$. Finally, the claim follows from Theorem 20.2. ❑

A finitely generated abelian group G is isomorphic to $G_{\text{tors}} \times \mathbb{Z}^r$ with some $r \in \mathbb{Z}_{\geq 0}$; here the torsion subgroup G_{tors} is finite (we already know this when $G = E(\mathbb{Q})$).

21.5. Definition. Let E be an elliptic curve over $\mathbb{Q}$ such that

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \times \mathbb{Z}^r \,.$$

Then $\text{rk}(E(\mathbb{Q})) := r \in \mathbb{Z}_{\geq 0}$ is the *rank of* $E(\mathbb{Q})$. ♢

DEF
rank

21.6. Lemma. *Let E be an elliptic curve over $\mathbb{Q}$ with rank r. Then*

$$r = \dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim_{\mathbb{F}_2} E(\mathbb{Q})[2] \,.$$

LEMMA
rank and
$E(\mathbb{Q})/2E(\mathbb{Q})$

Proof. We have

$$\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \cong \frac{E(\mathbb{Q})_{\text{tors}}}{2E(\mathbb{Q})_{\text{tors}}} \times \left(\frac{\mathbb{Z}}{2\mathbb{Z}}\right)^r .$$

Let $\mu \colon E(\mathbb{Q})_{\text{tors}} \to E(\mathbb{Q})_{\text{tors}}$, $T \mapsto 2T$. Then

$$2E(\mathbb{Q})_{\text{tors}} = \text{im}(\mu) \cong \frac{E(\mathbb{Q})_{\text{tors}}}{\ker(\mu)} = \frac{E(\mathbb{Q})_{\text{tors}}}{E(\mathbb{Q})[2]} \,,$$

hence $\#E(\mathbb{Q})[2] = \#(E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}})$. The claim now follows from

$$\dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} = \dim_{\mathbb{F}_2} \frac{E(\mathbb{Q})_{\text{tors}}}{2E(\mathbb{Q})_{\text{tors}}} + r = \dim_{\mathbb{F}_2} E(\mathbb{Q})[2] + r \,.$$ ❑

21.7. Definition. Let $n \in \mathbb{Z} \setminus \{0\}$. We write $\omega(n)$ for the number of distinct prime divisors of n. ♢

DEF
$\omega(n)$

21.8. Corollary. *Let $a, b \in \mathbb{Z}$ with $b, a^2 - 4b \neq 0$ and let*

$$E \colon y^2 = x(x^2 + ax + b) \,.$$

Then E is an elliptic curve over $\mathbb{Q}$, and we have that

$$\text{rk}(E(\mathbb{Q})) \leq \omega(b) + \omega(a^2 - 4b) \,.$$

COR
bound
for r

Proof. We see from the proof of Theorem 21.4 that

$$\dim \text{im}(\delta) \leq \dim H = 1 + \omega(b) \quad \text{and} \quad \dim \text{im}(\delta') \leq \dim H' = 1 + \omega(a^2 - 4b) \,.$$

Furthermore,

$$\dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} = \dim \text{im}(\delta) \qquad \text{and} \qquad \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} = \dim \text{im}(\delta') \,.$$

We have that $\dim E(\mathbb{Q})[2] \in \{1, 2\}$ (always ≤ 2 and > 0 since we have a point T of order 2). In the case that $\dim E(\mathbb{Q})[2] = 2$ we obtain

$$\begin{aligned} \text{rk}(E(\mathbb{Q})) &= \dim \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim E(\mathbb{Q})[2] \\ &\leq \dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} + \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 2 \\ &\leq (1 + \omega(b)) + (1 + \omega(a^2 - 4b)) - 2 = \omega(b) + \omega(a^2 - 4b) \,. \end{aligned}$$

In the other case $\dim E(\mathbb{Q})[2] = 1$ we have that $T' \notin \phi(E(\mathbb{Q}))$ since the preimages of T' under ϕ are exactly the two other points ($\neq T$) of order 2 on E. The coset

$T' + \phi(E(\mathbb{Q}))$ is then a nontrivial element of the kernel of the map β that we have considered in the proof of Lemma 21.1. It follows that

$$\dim \ker(\alpha) = \dim \operatorname{im}(\beta) \le \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 1\,.$$

Then we have again that (with α, β as in the proof of Lemma 21.1)

$$\begin{aligned} \operatorname{rk}(E(\mathbb{Q})) &= \dim \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} - \dim E(\mathbb{Q})[2] \\ &= \dim \operatorname{im}(\alpha) + \dim \ker(\alpha) - 1 \\ &\le \dim \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} + \dim \frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} - 2 \\ &\le \omega(b) + \omega(a^2 - 4b)\,. \end{aligned}$$

❑

The bound that we have just deduced is unfortunately never tight. In the next section we will show how it can be improved.

REM
The general case

21.9. **Remark.** What can we do when E has no rational point of order 2? Let $T = (\xi, 0) \in E[2]$ (we can assume that the coefficients a_1 and a_3 in the equation of E are zero), then $K = \mathbb{Q}(\xi)$ is a cubic field extension of $\mathbb{Q}$ and $T \in E(K)[2]$. We can then consider E as an elliptic curve over K and show by essentially the same proof as above that $E(K)/2E(K)$ is finite. This requires two basic results from Algebraic Number Theory (the finiteness of the class number and the finite generation of the group of units), which can be used to show that there are again finite subgroups H and H' of $K^\times/K^{\times 2}$ that contain $\operatorname{im}(\delta)$ and $\operatorname{im}(\delta')$, respectively. It then remains to show that the natural map $E(\mathbb{Q})/2E(\mathbb{Q}) \to E(K)/2E(K)$ has finite kernel. (This map need not be injective, since it is possible that a point $P \in E(\mathbb{Q})$ that is not divisible by 2 in $E(\mathbb{Q})$ becomes divisible by 2 in $E(K)$.) This can be seen as follows.

By potentially enlarging K, we can assume that K is a finite Galois extension of $\mathbb{Q}$ (if $E\colon y^2 = f(x)$, then we can take the splitting field of f). The kernel of the map under consideration can only grow when we do this. Let $\phi\colon E(\mathbb{Q}) \to E(K)/2E(K)$ be the canonical map and consider $P \in \ker(\phi)$. Then there is $Q \in E(K)$ with $P = 2Q$. We define a map $\delta_P\colon \operatorname{Gal}(K/\mathbb{Q}) \to E[2]$ by $\delta_P(\sigma) = \sigma(Q) - Q$. Then

$$2(\sigma(Q) - Q) = \sigma(2Q) - 2Q = \sigma(P) - P = O$$

(since $P \in E(\mathbb{Q})$), hence $\delta_P(\sigma) \in E[2]$. The map δ_P depends on the choice of Q; we fix one suitable Q for each P. In this way, we obtain a map

$$\delta\colon \ker(\phi) \longrightarrow \operatorname{Map}(\operatorname{Gal}(K/\mathbb{Q}), E[2])\,, \quad P \longmapsto \delta_P\,.$$

If $\delta_P = \delta_{P'}$, then we have for all $\sigma \in \operatorname{Gal}(K/\mathbb{Q})$ that

$$\sigma(Q) - Q = \sigma(Q') - Q' \implies \sigma(Q - Q') = Q - Q'\,;$$

this means that $R = Q - Q' \in E(\mathbb{Q})$. This implies that

$$P = 2Q = 2(Q' + R) = 2Q' + 2R = P' + 2R\,,$$

so $P + 2E(\mathbb{Q}) = P' + 2E(\mathbb{Q})$. This shows that representatives of distinct elements of the kernel of $E(\mathbb{Q})/2E(\mathbb{Q}) \to E(K)/2E(K)$ are mapped under δ to distinct maps $\operatorname{Gal}(K/\mathbb{Q}) \to E[2]$. Hence

$$\#\ker\Bigl(\frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \to \frac{E(K)}{2E(K)}\Bigr) \le \#\operatorname{Map}\bigl(\operatorname{Gal}(K/\mathbb{Q}), E[2]\bigr) = 4^{[K:\mathbb{Q}]} < \infty\,.$$

We can also use $\delta\colon E(\mathbb{Q}) \to K^\times/K^{\times 2}$, $P \mapsto (x(P) - \xi)K^{\times 2}$ (with $\delta(O) = K$), directly. One can show that $\ker(\delta) = 2E(\mathbb{Q})$, so that bounding the image of δ by a finite group immediately shows that $E(\mathbb{Q})/2E(\mathbb{Q})$ is finite.

Alternatively, we can generalize the definition of the height function h in such a way that it also applies to $E(K)$. Then one can show directly that $E(K)$ is finitely generated, which implies the claim also for $E(\mathbb{Q})$ (since $E(\mathbb{Q}) \subset E(K)$ is a subgroup). The difficulty with this approach is that there is no longer an essentially unique representation of the x-coordinate as a "fraction in lowest terms".

A different approach that bypasses the need for results from Algebraic Number Theory proceeds by defining an injective map δ on $E(\mathbb{Q})/2E(\mathbb{Q})$ with values in equivalence classes of homogeneous polynomials of degree 4 over $\mathbb{Q}$ in two variables (with respect to a suitable equivalence relation). One then shows that only finitely many of the equivalence classes can occur as images under δ. This requires a detailed study of the so-called invariant theory of these quartics. See [Cre]. ♠

22. A better bound for the rank

From the considerations in the previous section we can conclude that we can determine the rank of an elliptic curve E over $\mathbb{Q}$ such that $E(\mathbb{Q})[2] \neq \{O\}$, if we can find the images of δ and δ': the proof of Corollary 21.8 shows that

$$\operatorname{rk}(E(\mathbb{Q})) = \dim \operatorname{im}(\delta) + \dim \operatorname{im}(\delta') - 2\,. \tag{22.1}$$

The problem here is that it is not so easy to determine these images exactly. (Indeed, there is no method known so far that could be shown to that in all cases.)

As usual, let

$$E\colon y^2 = x(x^2 + ax + b)$$

with $a, b \in \mathbb{Z}$ and $b, a^2 - 4b \neq 0$. For simplicity we will write d in the following when we really mean $d\mathbb{Q}^{\times 2}$.

We know that $\operatorname{im}(\delta) \subset H$, where H is the subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ that is generated by (the square classes of) -1 and the prime divisors of b. Hence the task of determining $\operatorname{im}(\delta)$ is equivalent to deciding for each element $d \in H$ whether there is some $P \in E(\mathbb{Q})$ such that $\delta(P) = d$, i.e., such that $x(P) = dt^2$ for some $t \in \mathbb{Q}^\times$ (we exclude the cases $P = O$ and $P = T$ here, as we know their images 1 and b under δ). If we substitute this into the equation of E, we obtain

$$y^2 = dt^2(d^2t^4 + adt^2 + b)\,,$$

or equivalently,

$$\left(\frac{y}{dt}\right)^2 = dt^4 + at^2 + \frac{b}{d}\,.$$

If we choose d to be a squarefree integer, then $b/d \in \mathbb{Z}$ since d is up to sign a product of prime divisors of b. Writing $t = u/v$ with coprime integers u, v and setting $w = yv^3/(du)$, we obtain

$$w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4\,. \tag{22.2}$$

Every point $P \in E(\mathbb{Q})$ with $\delta(P) = d$ therefore leads to a solution of (22.2) in integers u, v, w with $u \perp v$. This remains true when $P = O$ or $P = T$. In the first case, we have the solution $(u, v, w) = (1, 0, 1)$ with $d = 1$ and in the second case, $(u, v, w) = (0, 1, 1)$ with $d = b$.

The existence of an integral solution with $u \perp v$ is equivalent to the existence of a rational solution with $(u, v) \neq (0, 0)$ since when (u, v, w) is a solution, then so is $(\lambda u, \lambda v, \lambda^2 w)$. If we replace d by $s^2 d$ with some $s \in \mathbb{Q}^\times$, then we obtain an equivalent equation by scaling u and v by s. This shows that the solubility of Equation (22.2) indeed depends only on the square class of d.

Conversely, if we have a solution (u, v, w) of (22.2) with $(u, v) \neq (0, 0)$, then

$$P = \left(d\frac{u^2}{v^2}, d\frac{uw}{v^3}\right) \in E(\mathbb{Q})$$

(if $v = 0$, we take $P = O$) such that $\delta(P) = d$.

How can we decide whether (22.2) has a solution? If there is a solution, then we can find it (at least in principle). Verifying that there is no solution is usually harder. One possibility is to consider the equation modulo n for a suitable $n \geq 2$. If the congruence

$$w^2 \equiv du^4 + au^2v^2 + \frac{b}{d}v^4 \bmod n$$

has no solution in $\mathbb{Z}$ such that $\gcd(u, v, n) = 1$, then (22.2) has no nontrivial solution. The Chinese Remainder Theorem implies that it suffices to consider the case that n is a prime power. It is also possible that the right hand side of the equation is always negative; then again no solution can exist.

EX Determination of $E(\mathbb{Q})$

22.1. **Example.** We consider the elliptic curve

$$E\colon y^2 = x^3 + x = x(x^2 + 1)\,.$$

The isogenous curve is

$$E'\colon y^2 = x(x^2 - 4)\,.$$

We have the following bounds for the images of δ and of δ'.

$$\operatorname{im}(\delta) \subset H = \langle -1 \rangle\,, \qquad \operatorname{im}(\delta') \subset H' = \langle -1, 2 \rangle\,.$$

For $-1 \in H$ we obtain from (22.2) the equation

$$w^2 = -u^4 - v^4$$

that does not even have a real solution with $(u, v) \neq (0, 0)$. This implies that $-1 \notin \operatorname{im}(\delta)$ and hence that $\operatorname{im}(\delta) = \{1\}$. Using (22.1) we obtain

$$0 \le \operatorname{rk}(E(\mathbb{Q})) = \dim \operatorname{im}(\delta) + \dim \operatorname{im}(\delta') - 2 \le 0 + 2 - 2 = 0\,,$$

so $\operatorname{rk}(E(\mathbb{Q})) = 0$. Therefore,

$$E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} = \{O, (0, 0)\}\,;$$

the second equality follows for example from the Nagell-Lutz Theorem 18.5.

(We also see that we must have $\operatorname{im}(\delta') = H'$. Indeed,

$$\delta'(0, 0) = -1\,, \qquad \delta'(2, 0) = 2\,, \qquad \delta'(-2, 0) = -2\,.) \qquad ♣$$

EX Determination of $\operatorname{rk}(E(\mathbb{Q}))$

22.2. **Example.** This time, we consider

$$E\colon y^2 = x(x^2 + 10x + 8)\,, \qquad E'\colon y^2 = x(x^2 - 20x + 68)\,.$$

We have

$$\operatorname{im}(\delta) \subset H = \langle -1, 2 \rangle \qquad \text{and} \qquad \operatorname{im}(\delta') \subset H' = \langle -1, 2, 17 \rangle\,.$$

We know that $\langle 2 \rangle \subset \operatorname{im}(\delta)$, for $\delta(T) = 8 = 2$ (modulo squares). It therefore suffices to consider $d = -1$ (say). This gives the equation

$$w^2 = -u^4 + 10u^2v^2 - 8v^4\,,$$

which has the solution $(u, v, w) = (1, 1, 1)$. So $-1 = \delta(-1, 1) \in \operatorname{im}(\delta)$ and hence $\operatorname{im}(\delta) = H$.

On the other hand, we have $\langle 17 \rangle \subset \operatorname{im}(\delta')$ since $\delta'(T') = 68 = 17$. We consider $d = -1$; this gives

$$w^2 = -u^4 - 20u^2v^2 - 68v^4\,.$$

The right hand side is always negative, so there is no solution. The same argument works for every $d < 0$; this shows that

$$\operatorname{im}(\delta') \subset \langle 2, 17 \rangle\,.$$

Now we consider $d = 2$. The equation is

$$w^2 = 2u^4 - 20u^2v^2 + 34v^4\,.$$

It has the solution $(u, v, w) = (1, 1, 4)$. We see that $\operatorname{im}(\delta') = \langle 2, 17 \rangle$, which finally implies that

$$\operatorname{rk}(E(\mathbb{Q})) = \dim \operatorname{im}(\delta) + \dim \operatorname{im}(\delta') - 2 = 2 + 2 - 2 = 2\,. \qquad ♣$$

We see in these examples that it is frequently possible to exclude potential values of d because they lead to negative definite quartics on the right hand side of (22.2). The following lemma provides criteria for when this is the case.

LEMMA
negative definite quartic

22.3. **Lemma.** *Let $E\colon y^2 = x(x^2 + ax + b)$ with $a, b \in \mathbb{Z}$ be an elliptic curve. If $a^2 - 4b < 0$ or ($a \le 0$ and $b > 0$), then*

$$\operatorname{im}(\delta) \subset \langle p : p \text{ prime}, p \mid b \rangle .$$

Proof. The claim is equivalent to the statement that negative d cannot occur as values of δ. So let $d = -|d| < 0$. The quartic on the right hand side of (22.2) then is

$$-\bigl(|d|u^4 - au^2v^2 + (b/|d|)v^4\bigr) .$$

The discriminant of the quadratic form $|d|X^2 - aXY + (b/|d|)Y^2$ is $a^2 - 4|d|(b/|d|) = a^2 - 4b$. If it is negative, then the expression in parentheses above can take only positive values (for $(u, v) \ne (0, 0)$); therefore the equation has no nontrivial real solution. If $a \le 0$ and $b > 0$, then all terms in the expression are ≥ 0 (and all $= 0$ only when $u = v = 0$), so there is again no nontrivial real solution. ❑

The converse is also true: if $a^2 - 4b > 0$ and ($a > 0$ or $b < 0$), then the equation for $d < 0$ does have a nontrivial real solution (exercise). In this sense, the statement of the lemma is best possible.

One can use Lemma 22.3 to improve the bound from Corollary 21.8 as follows (exercise).

$$\operatorname{rk}(E(\mathbb{Q})) \le \omega(b) + \omega(a^2 - 4b) - 1 .$$

Examples 22.1 and 22.2 show that this bound can be tight. That this is not true in general is shown by the following example.

EX
Determination of $\operatorname{rk}(E(\mathbb{Q}))$

22.4. **Example.** Let

$$E\colon y^2 = x(x^2 - 15x + 63) \qquad \text{and} \qquad E'\colon y^2 = x(x^2 + 30x - 27) .$$

We have $a = -15 < 0$ and $b = 63 > 0$; by Lemma 22.3 we therefore obtain that

$$\operatorname{im}(\delta) \subset \langle 3, 7\rangle \qquad \text{and} \qquad \operatorname{im}(\delta') \subset \langle -1, 3\rangle .$$

Furthermore, $\langle 7\rangle \subset \operatorname{im}(\delta)$ and $\langle -3\rangle \subset \operatorname{im}(\delta')$. We try to decide if $3 \in \operatorname{im}(\delta)$. The relevant equation is

$$w^2 = 3u^4 - 15u^2v^2 + 21v^4$$

with the solution $(u, v, w) = (1, 1, 3)$. So $\operatorname{im}(\delta) = \langle 3, 7\rangle$. Next we consider $d = 3$ for δ'. The equation is

$$w^2 = 3u^4 + 30u^2v^2 - 9v^4 .$$

The right hand side is divisible by 3, so $w = 3w_1$ with $w_1 \in \mathbb{Z}$. We obtain the new equation

$$3w_1^2 = u^4 + 10u^2v^2 - 3v^4 \equiv u^2(u^2 + v^2) \bmod 3 .$$

Since $u^2 + v^2$ is divisible by 3 only when u and v are both divisible by 3, which contradicts the condition $u \perp v$, we must have that $u = 3u_1$ with $u_1 \in \mathbb{Z}$ (and $3 \nmid v$). Plugging this in leads to

$$w_1^2 = 27u_1^4 + 30u_1^2v^2 - v^4 \equiv -v^4 \bmod 3 .$$

This is only possible when v is divisible by 3, a contradiction. So the equation has no solution, and $\operatorname{im}(\delta') = \langle -3\rangle$. Altogether, we see that

$$\operatorname{rk}(E(\mathbb{Q})) = \dim \operatorname{im}(\delta) + \dim \operatorname{im}(\delta') - 2 = 2 + 1 - 2 = 1 . \qquad ♣$$

These examples motivate the following definitions.

22.5. **Definition.** We say that an equation of the form (22.2) is *everywhere locally soluble (ELS)*, if it has nontrivial solutions in $\mathbb{R}$ and nontrivial solutions modulo n for all $n \in \mathbb{Z}_{\geq 2}$. ♢

DEF everywhere locally soluble

"Nontrivial" over a field means that $(u, v) \neq (0, 0)$, and "nontrivial" modulo n means that $\gcd(u, v, n) = 1$. It is clear that a nontrivially soluble equation is also everywhere locally soluble.

22.6. **Definition.** Let E, E' and $\hat{\phi}$ as usual. Then

$$S_{\hat{\phi}} := \{d \in \mathbb{Q}^\times/\mathbb{Q}^{\times 2} \mid (22.2) \text{ is ELS}\}$$

is the *Selmer group* of the isogeny $\hat{\phi}$. ♢

DEF Selmer group

The set $S_{\hat{\phi}}$ indeed is a subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$.

One can see this as follows. The everywhere local solubility is equivalent to the nontrivial solubility in $\mathbb{R}$ and in the field $\mathbb{Q}_p$ of p-adic numbers for every prime p. Over every field $K \supset \mathbb{Q}$, it is still true (with the same proof as over $\mathbb{Q}$) that the equation for d has a nontrivial solution in K if and only if there is $P \in E(K)$ with $\delta_K(P) = d$ (where $\delta_K \colon E(K) \to K^\times/K^{\times 2}$ is defined in the same way as δ). If the equations for d and for d' both have nontrivial solutions in K, then there are $P, P' \in E(K)$ with $\delta_K(P) = d$ and $\delta_K(P') = d'$. Since δ_K is a homomorphism, we have that $\delta_K(P + P') = dd'$, which implies in turn that the equation for dd' also has a nontrivial solution in K. Applying this to $K = \mathbb{R}$ and to $K = \mathbb{Q}_p$ for all primes p, we see that $S_{\hat{\phi}}$ is a group.

22.7. **Lemma.** *With notations introduced earlier we have $S_{\hat{\phi}} \subset H$.*

LEMMA $S_{\hat{\phi}} \subset H$

Proof. Assume that $d \notin H$. Then there is a prime p such that $p \nmid b$ and $p \mid d$; we write $d = pd'$ with $p \nmid d'$. The equation for d is equivalent with

$$w^2 = d^3u^4 + ad^2u^2v^2 + bdv^4 = p^3{d'}^3u^4 + p^2a{d'}^2u^2v^2 + pbd'v^4 .$$

We then must have $w = pw_1$ with $w_1 \in \mathbb{Z}$. This leads to

$$pw_1^2 = p^2{d'}^3u^4 + pa{d'}^2u^2v^2 + bd'v^4 .$$

Since $p \nmid bd'$, we must then have $v = pv_1$ with $v_1 \in \mathbb{Z}$. This now leads to

$$w_1^2 = p{d'}^3u^4 + p^2a{d'}^2u^2v_1^2 + p^3bd'v_1^4$$

and then to $w_1 = pw_2$ with $w_2 \in \mathbb{Z}$, so

$$pw_2^2 = {d'}^3u^4 + pa{d'}^2u^2v_1^2 + p^2bd'v_1^4 .$$

But then u must also be divisible by p, contradicting $u \perp v$. This shows that there is no nontrivial solution modulo p^4 and so $d \notin S_{\hat{\phi}}$. ❑

We now show how to get a bound on the rank from $S_{\hat{\phi}}$ and S_ϕ.

22.8. **Lemma.** *In the situation of the definition above we have that*

LEMMA bound for $\operatorname{im}(\delta)$

$$\frac{E(\mathbb{Q})}{\hat\phi(E'(\mathbb{Q}))} \cong \operatorname{im}(\delta) \subset S_{\hat\phi}\,.$$

In particular, we have the estimate

$$\operatorname{rk}(E(\mathbb{Q})) \le \dim S_{\hat\phi} + \dim S_{\phi} - 2\,.$$

Proof. We already had seen that there is an isomorphism $E(\mathbb{Q})/\hat\phi(E'(\mathbb{Q})) \cong \operatorname{im}(\delta)$. If $d \in \operatorname{im}(\delta)$, then the equation for d has an integral solution with $u \perp v$. This solution then is also a nontrivial real solution and a nontrivial solution mod n for all $n \ge 2$. Hence $d \in S_{\hat\phi}$. The second claim then follows from (22.1). ❑

22.9. **Remark.** More generally, one can define for every isogeny $\varphi\colon E' \to E$ of elliptic curves over $\mathbb{Q}$ (or even more generally, over an algebraic number field) a Selmer group S_φ together with a map $\delta\colon E(\mathbb{Q}) \to S_\varphi$ such that $\ker(\delta) = \varphi(E'(\mathbb{Q}))$. The Selmer group S_φ is finite and can be computed (at least in principle). (We will soon show this when ϕ is an isogeny of degree 2.) If $\varphi = [m]$ is the multiplication by $m \ge 2$, then one directly obtains a bound on $\operatorname{rk}(E(\mathbb{Q}))$, namely,

REM Selmer groups

$$\operatorname{rk}(E(\mathbb{Q})) \le \log_m \#\frac{S_{[m]}}{\delta(E(\mathbb{Q})_{\text{tors}})}\,.$$

Otherwise one needs (bounds for) $\#S_\varphi$ and $\#S_{\hat\varphi}$, in analogy with Lemma 22.8. In practice, one can compute $S_{[m]}$ for $m = 2, 3, 4$ and, with restrictions, $m = 8, 9$. For isogenies φ one can compute S_φ when $\deg\varphi$ is reasonably small.

To set up the general definition of the Selmer group S_φ we first define for each field extension K of $\mathbb{Q}$

$$\begin{aligned} Z^1(K, \ker(\varphi)) &= \{\xi\colon \operatorname{Gal}_K \to \ker(\varphi) \mid \forall \sigma, \tau \in \operatorname{Gal}_K\colon \xi(\sigma\tau) = \sigma(\xi(\tau)) - \xi(\sigma)\}\,,\\ B^1(K, \ker(\varphi)) &= \{\xi\colon \operatorname{Gal}_K \to \ker(\varphi),\ \sigma \mapsto \sigma(T) - T \mid T \in \ker(\varphi)\} \qquad \text{and}\\ H^1(K, \ker(\varphi)) &= Z^1(K, \ker(\varphi))/B^1(K, \ker(\varphi))\,. \end{aligned}$$

The elements of $Z^1(K, \ker(\varphi))$ are called *1-cocycles on* Gal_K *with values in* $\ker(\varphi)$, the elements of $B^1(K, \ker(\varphi))$ are called *1-coboundaries on* Gal_K *with values in* $\ker(\varphi)$, and $H^1(K, \ker(\varphi))$ is the *first Galois cohomology group over* K *with values in* $\ker(\varphi)$. Here $\operatorname{Gal}_K = \operatorname{Aut}(\bar K/K)$ is the absolute Galois group of K. $Z^1(K, \ker(\varphi))$ is a subgroup of $\operatorname{Map}(\operatorname{Gal}_K, \ker(\varphi))$ and $B^1(K, \ker(\varphi))$ is a subgroup of $Z^1(K, \ker(\varphi))$, so that $H^1(K, \ker(\varphi))$ is well-defined as a group. The map δ is then defined by

$$\delta\colon E(\mathbb{Q}) \longrightarrow H^1(\mathbb{Q}, \ker(\varphi))\,, \qquad P \longmapsto [\sigma \mapsto \sigma(Q) - Q]\,,$$

where $Q \in E'(\bar{\mathbb{Q}})$ is such that $\varphi(Q) = P$ and $[\xi]$ denotes the coset of ξ in $H^1(\mathbb{Q}, \ker(\varphi))$. One easily checks that δ is well-defined (i.e., the map $\sigma \mapsto \sigma(Q) - Q$ is in $Z^1(\mathbb{Q}, \ker(\varphi))$, and this cocycle changes by a coboundary when Q is replaced by a different preimage of P). For each $v = p$ prime or $v = \infty$ one has a commutative diagram (where $\mathbb{Q}_\infty := \mathbb{R}$)

$$\begin{tikzcd} E(\mathbb{Q}) \arrow[r, "\delta"] \arrow[d, hook] & H^1(\mathbb{Q}, \ker(\varphi)) \arrow[d, "r_v"] \\ E(\mathbb{Q}_v) \arrow[r, "\delta_v"] & H^1(\mathbb{Q}_v, \ker(\varphi)) \end{tikzcd}$$

and one defines

$$S_\varphi := \{\xi \in H^1(\mathbb{Q}, \ker(\varphi)) \mid \forall v\colon r_v(\xi) \in \operatorname{im}(\delta_v)\}\,.$$

Then $\operatorname{im}(\delta) \subset S_\varphi$.

The connection with our definition of $S_{\hat\phi}$ is as follows. The action of Gal_K on $\ker(\hat\phi)$ is trivial (for every field $K \supset \mathbb{Q}$). This implies that

$$H^1(K, \ker(\hat\phi)) \cong \mathrm{Hom}(\mathrm{Gal}_K, \ker(\hat\phi))\,.$$

On the other hand, one can show that

$$\frac{K^\times}{K^{\times 2}} \longrightarrow \mathrm{Hom}(\mathrm{Gal}_K, \{\pm 1\})\,, \qquad \alpha K^{\times 2} \longmapsto \Bigl(\sigma \mapsto \frac{\sigma(\sqrt{\alpha})}{\sqrt{\alpha}}\Bigr)$$

is an isomorphism. As a "Galois module" (i.e., an abelian group with an action of Gal_K), $\ker(\hat\phi)$ is isomorphic to $\{\pm 1\}$, so that we obtain our δ as the composition of the first isomorphism with the inverse of the second. In this way, we can define $S_{\hat\phi}$ as a subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$, and the condition "$r_v(\xi) \in \mathrm{im}(\delta_v)$" above is equivalent to the existence of nontrivial solutions in $\mathbb{R}$ and in $\mathbb{Q}_p$ for all primes p of the equation belonging to $\xi \leftrightarrow d\mathbb{Q}^{\times 2}$. ♠

We now want to figure out how one can determine $S_{\hat\phi}$ algorithmically. Since we already know that $S_{\hat\phi} \subset H$ with an explicit finite group H, this task is equivalent with determining, for a given $d \in H$, whether the equation

$$w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4$$

is ELS. Lemma 22.3 gives a criterion for the existence of a nontrivial solution in $\mathbb{R}$ when $d < 0$ (if $d > 0$, there is always the solution $(u, v, w) = (1, 0, \sqrt{d})$). It therefore remains to decide whether the equation has a nontrivial solution mod p^e for all primes p and all $e \geq 1$.

22.10. **Lemma.** *Let $a, c, d \in \mathbb{Z}$ and let p be an odd prime such that $p \nmid cd$ and $p \nmid a^2 - 4cd$. Then the equation*

$$w^2 = du^4 + au^2v^2 + cv^4$$

has nontrivial solutions mod p^e for all $e \geq 1$.

LEMMA
solubility
for almost
all p

Proof. We write $\bar n$ for the image of $n \in \mathbb{Z}$ in $\mathbb{F}_p$. Let $E\colon y^2 = x(x^2 + \bar a x + \bar c \bar d)$; E is an elliptic curve over $\mathbb{F}_p$. There is an isogeny $\hat\phi\colon E' \to E$ of degree 2. We can define the map $\delta\colon E(\mathbb{F}_p) \to \mathbb{F}_p^\times/\mathbb{F}_p^{\times 2}$ in the same way as over $\mathbb{Q}$, and one also sees in the same way that the existence of a nontrivial solution mod p of the equation in the statement of the lemma is equivalent to the existence of a point $P \in E(\mathbb{F}_p)$ such that $\delta(P) = \bar d \mathbb{F}_p^{\times 2}$. Similarly, $\ker(\delta) = \hat\phi(E'(\mathbb{F}_p))$. The claim for $e = 1$ then is equivalent to δ being surjective. Now $E'(\mathbb{F}_p)$ and $E(\mathbb{F}_p)$ are finite abelian groups. By (the easy direction in) Theorem 13.3, they have the same order. We see that

$$\begin{aligned} \#\,\mathrm{im}(\delta) &= \bigl(E(\mathbb{F}_p) : \ker(\delta)\bigr) = \bigl(E(\mathbb{F}_p) : \hat\phi(E'(\mathbb{F}_p))\bigr) \\ &= \frac{\#E(\mathbb{F}_p)}{\#\hat\phi(E'(\mathbb{F}_p))} = \frac{\#E(\mathbb{F}_p)}{\#E'(\mathbb{F}_p)/\#\ker(\hat\phi)} = \#\ker(\hat\phi) = 2 = \#\frac{\mathbb{F}_p^\times}{\mathbb{F}_p^{\times 2}}\,, \end{aligned}$$

and so δ is surjective.

So there is a nontrivial solution (u, v, w) mod p. We can scale u and v in such a way that $\bar u = 1$ or $\bar v = 1$. Since only the residue classes of u, v, w mod p are relevant, we can therefore assume that $u = 1$ or $v = 1$. Without loss of generality, assume that $v = 1$ (the other case is symmetric). We then have

$$w^2 \equiv du^4 + au^2 + c \bmod p\,.$$

If p does not divide w, then the image of the right hand side in $\mathbb{F}_p$ is a nonzero square. Hensel's Lemma 22.11 below then implies that the right hand side is a square mod p^e for all $e \geq 1$; this shows the claim when $p \nmid w$.

If $p \mid w$, then $\bar{u}$ is a root of $\bar{d}x^4+\bar{a}x^2+\bar{c}$. This root is simple since p does not divide the discriminant $16cd(a^2-4cd)^2$ of dx^4+ax^2+c. Again by Hensel's Lemma 22.11 it follows that for each $e \geq 1$ there is $u_e \in \mathbb{Z}$ such that $du_e^4+au_e^2+c \equiv 0 \bmod p^e$. This shows the claim also when $p \mid w$. ❑

In the proof above we have used the following fact.

22.11. **Theorem.** *Let $f \in \mathbb{Z}[x]$ be a polynomial and p be a prime. Let further $u \in \mathbb{Z}$ such that $p \mid f(u)$ and $p \nmid f'(u)$. Then for every $e \geq 1$ there exists an integer u_e such that $u_e \equiv u \bmod p$ and $p^e \mid f(u_e)$. The residue class of u_e mod p^e is uniquely determined.*

THM
Hensel's
Lemma

Proof. We show the claim by induction on e. The base case $e=1$ is given by the assumptions. For the inductive step, assume that we have $u_e \in \mathbb{Z}$ such that $u_e \equiv u \bmod p$ and $f(u_e)=\alpha p^e$ for some $\alpha \in \mathbb{Z}$. The uniqueness part of the inductive hypothesis shows that we must have $u_{e+1} \equiv u_e \bmod p^e$, so we set $u_{e+1}=u_e+xp^e$ with x to be determined. We obtain

$$f(u_{e+1})=f(u_e+xp^e) \equiv f(u_e)+f'(u_e)xp^e=(\alpha+f'(u_e)x)p^e \bmod p^{e+1}\,.$$

We have $f'(u_e) \equiv f'(u) \not\equiv 0 \bmod p$, hence the congruence

$$\alpha+f'(u_e)x \equiv 0 \bmod p$$

has a solution x that is uniquely determined mod p. This shows the existence of u_{e+1} and the uniqueness mod p^{e+1}. ❑

22.12. **Corollary.** *Let $f \in \mathbb{Z}[x]$ be a polynomial and p be a prime. Let further $u \in \mathbb{Z}$ and set $e_0=v_p(f'(u))$. If $v_p(f(u))>2e_0$, then for each $e>2e_0$ there exists an integer u_e such that $u_e \equiv u \bmod p^{e_0+1}$ and $p^e \mid f(u_e)$.*

COR
Hensel's
Lemma
(variant)

Proof. Set $F(x)=p^{-2e_0}f(u+p^{e_0}x) \in \mathbb{Z}[x]$. The claim follows from Theorem 22.11 with $(f,p,u) \leftarrow (F,p,0)$. ❑

Lemma 22.10 reduces the problem to the consideration of finitely many primes p, namely, $p=2$, the prime divisors of b, and the prime divisors of a^2-4b. It remains to show that one can decide the existence of nontrivial solutions of our equation modulo all powers of p for a *single* prime p. Hensel's Lemma will again play a decisive role.

22.13. **Lemma.** *Let $a,c,d \in \mathbb{Z}$ with $cd \neq 0$ and $a^2-4cd \neq 0$. Let further p be a prime and set $e_0=2v_p(4cd(a^2-4cd))+1$. The equation*

LEMMA
solubility
mod p^e

$$w^2=du^4+au^2v^2+cv^4$$

has nontrivial solutions mod p^e for all $e \geq 1$ if and only if it has a nontrivial solution mod p^{e_0}.

Proof. The direction "⇒" is trivial. To show "⇐", we can as in the proof of Lemma 22.10 assume without loss of generality that our nontrivial solution mod p^{e_0} has the form $(u,1,w)$. Set $f(x)=dx^4+ax^2+c$. Then $v_p(f(u))$ must be either $< e_0$ and even or $\geq e_0$.

First assume that p is odd. If $v_p(f(u)) < e_0$, then $p^{-v_p(f(u))}f(u)$ is a quadratic residue mod p; this again implies (by Hensel's Lemma) that $f(u)$ is a square mod p^e for all $e \geq 1$. In the case $v_p(f(u)) \geq e_0$ we consider the relation

$$(4adu^2 + 2(a^2 - 4cd))f(u) - (adu^3 + (a^2 - 2cd)u)f'(u) = 2c(a^2 - 4cd)\,.$$

If we had $v_p(f'(u)) \geq e_0/2$, then using that $v_p(f(u)) \geq e_0$ this would imply

$$v_p(2cd(a^2 - 4cd)) \geq e_0/2\,,$$

contradicting the definition of e_0. So we must have $v_p(f'(u)) < e_0/2$ and hence $v_p(f(u)) > 2v_p(f'(u))$. According to Corollary 22.12, for each e there exists $u_e \in \mathbb{Z}$ such that $f(u_e) \equiv 0 \bmod p^e$; then $(u, v, w) = (u_e, 1, 0)$ is a nontrivial solution mod p^e.

Now consider $p = 2$. It is still true that $v_2(f(u))$ is either $< e_0$ and even or $\geq e_0$. If $v_2(f(u)) \leq e_0 - 3$, then we must have $2^{-v_2(f(u))}f(u) \equiv 1 \bmod 8$; then $f(u)$ is a square mod 2^e for all $e \geq 1$. Otherwise $v_2(f(u)) \geq e_0 - 1$. Similarly as above, $v_2(f'(u)) \geq (e_0 - 1)/2$ would imply that $v_2(4cd(a^2 - cd)) > e_0/2$, contradicting the definition of e_0. In the same way as above, it follows that for every e there is some u_e such that $f(u_e) \equiv 0 \bmod 2^e$, which shows the existence of a nontrivial solutions. ❑

22.14. Corollary. *Let $\hat{\phi}\colon E' \to E$ be an isogeny of degree 2 between two elliptic curves over $\mathbb{Q}$. Then the Selmer group $S_{\hat{\phi}}$ is computable.*

COR computability of the Selmer group

Proof. We can assume that E has the form $y^2 = x(x^2 + ax + b)$ with $a, b \in \mathbb{Z}$ (then $b \neq 0$ and $a^2 - 4b \neq 0$). Let H be the finite subgroup of $\mathbb{Q}^\times/\mathbb{Q}^{\times 2}$ that is generated by -1 and the prime divisors of b. Then $S_{\hat{\phi}} \subset H$, so it suffices to check for each of the finitely many elements of H if it is contained in $S_{\hat{\phi}}$. The solubility of the relevant equation over $\mathbb{R}$ can easily be checked. Lemma 22.10 shows that we need to check the solubility mod p^e only for finitely many primes p, and Lemma 22.13 reduces this for each given prime to a finite problem. ❑

This leads to the following method for trying to determine the rank $\operatorname{rk}(E(\mathbb{Q}))$ when $E(\mathbb{Q})[2] \neq \{O\}$. Let $E\colon y^2 = x(x^2 + ax + b)$ with $a, b \in \mathbb{Z}$ as usual.

1. Determine the prime divisors of b and of $a^2 - 4b$ and thence H and H'.
2. For each $d \in H$, determine if $d \in S_{\hat{\phi}}$, and for each $d' \in H'$, determine if $d' \in S_\phi$.
3. For each $d \in S_{\hat{\phi}}$ and for each $d' \in S_\phi$, try to find a nontrivial integral solution of the associated equation. Let T and T' be the subsets of $S_{\hat{\phi}}$ and of S_ϕ, respectively, for which this is successful.
4. If $\langle T\rangle = S_{\hat{\phi}}$ and $\langle T'\rangle = S_\phi$, then

   $$\operatorname{rk}(E(\mathbb{Q})) = \dim S_{\hat{\phi}} + \dim S_\phi - 2\,.$$

   In any case,

   $$\dim\langle T\rangle + \dim\langle T'\rangle - 2 \leq \operatorname{rk}(E(\mathbb{Q})) \leq \dim S_{\hat{\phi}} + \dim S_\phi - 2\,.$$

There is, however, no guarantee that this method is successful. An equation of the form $w^2 = du^4 + au^2v^2 + cv^4$ that has nontrivial solutions in $\mathbb{R}$ and modulo n for all $n \geq 2$ need not have nontrivial solutions in $\mathbb{Z}$.

22.15. **Example.** The equation

EX
Selmer bound
not tight

$$w^2 = 2u^4 - 34v^4$$

obviously has nontrivial solutions in $\mathbb{R}$. It also has nontrivial solutions mod p^e for all primes p and exponents $e \geq 1$. For $p \neq 2, 17$ this follows from Lemma 22.10. For $p = 17$ we have the nontrivial solution $(u, v, w) = (1, 0, 6)$ mod 17, which by Hensel's Lemma can be lifted to a solution $(1, 0, w_e)$ mod 17^e for every $e \geq 1$. For $p = 2$, note that 17 is a fourth power modulo every power of 2: $17 \equiv 3^4 \bmod 2^5$, and the derivative $f'(x) = 4x^3$ of $f(x) = x^4 - 17$ satisfies $v_2(f'(3)) = 2$; the claim then follows from Corollary 22.12. So there is always a solution of the form $(u_e, 1, 0)$ mod 2^e.

On the other hand, there is *no* nontrivial integral solution. To see this, assume that (u, v, w) is an integral solution with $u \perp v$. Then certainly $w \neq 0$. Let p be an odd prime divisor of w. Then p cannot divide u or v, since otherwise p would have to divide both, which would contradict $u \perp v$. This also implies that $p \neq 17$, since otherwise $17 \mid u$ and $17 \mid v$. Then $u^4 \equiv 17v^4 \bmod p$ implies that 17 is a quadratic residue mod p. By the Law of Quadratic Reciprocity (use that $17 \equiv 1 \bmod 4$), p then is a quadratic residue mod 17. Since -1 and 2 are quadratic residues mod 17 as well, w, as a product of quadratic residues, is also a quadratic residue mod 17. So there is $t \in \mathbb{Z}$ such that $w \equiv t^2 \bmod 17$. This gives $t^4 \equiv 2u^4 \bmod 17$, which (since $2 \equiv x^4 \bmod 17$ has no solution) implies that $17 \mid t$ and $17 \mid u$. But then we also have $17 \mid v$, contradicting $u \perp v$. This shows that there is no nontrivial integral solution. (This example was first discovered independently by Lind[10] and Reichardt[11].)

The equation considered above belongs to $d = 2 \in S_{\hat{\phi}}$ for the pair

$$E\colon y^2 = x(x^2 - 68)\,, \qquad E'\colon y^2 = x(x^2 + 272)\,.$$

Here $H = H' = \langle -1, 2, 17 \rangle$, and we find that

$$S_{\hat{\phi}} = \langle -1, 2, 17 \rangle = H \qquad \text{and} \qquad S_{\phi} = \langle 17 \rangle\,,$$

leading to the bound $\operatorname{rk}(E(\mathbb{Q})) \leq 2$. But we have in fact that

$$\operatorname{im}(\delta) = \delta(\langle T \rangle) = \langle -17 \rangle$$

and hence $\operatorname{rk}(E(\mathbb{Q})) = 0$, so that $E(\mathbb{Q}) = E(\mathbb{Q})_{\text{tors}} = \{O, T\}$. The consideration above shows that $2 \notin \operatorname{im}(\delta)$. Similarly, one can show that $-2, 34, -34 \notin \operatorname{im}(\delta)$. Therefore $\operatorname{im}(\delta) \subset \langle -1, 17 \rangle$. We now show that $-1 \notin \operatorname{im}(\delta)$. The associated equation is

$$w^2 = -u^4 + 68v^4\,. \tag{22.3}$$

This implies that u and w must be even: $u = 2u_1$, $w = 2w_1$ and so

$$w_1^2 = -4u_1^4 + 17v^4$$

with v odd. Setting $U = (u_1/v)^2$, we then have $(w_1/v^2)^2 + (2U)^2 = 17$. We now set $w_1/v^2 = 1 - \lambda$ and $U = 2 - \lambda t$, which gives

$$\lambda(-2 + \lambda - 16t + 4\lambda t^2) = 0\,.$$

[10] Carl-Erik Lind: *Untersuchungen über die rationalen Punkte der ebenen kubischen Kurven vom Geschlecht Eins,* Uppsala: Diss. 97 S. (1940).

[11] Hans Reichardt: *Einige im Kleinen überall lösbare, im Grossen unlösbare diophantische Gleichungen,* J. reine angew. Math. **184** (1942), 12–18.

$\lambda = 0$ does not lead to a solution since $2 = U = (u_1/v)^2$ is impossible. Therefore

$$\lambda = \frac{4 + 16t}{1 + 4t^2},$$

which gives

$$\frac{w_1}{v^2} = \frac{-1 - 16t + 4t^2}{1 + 4t^2} \qquad \text{and} \qquad \left(\frac{u_1}{v}\right)^2 = U = \frac{2 - 2t - 8t^2}{1 + 4t^2}.$$

Here $t \in \mathbb{Q}$. (The limit case $t = \infty$ leads $U = -2$, which is also not a square.) If we write $t = r/s$ in lowest terms, then we obtain

$$\frac{u_1^2}{v^2} = \frac{-8r^2 - 2rs + 2s^2}{4r^2 + s^2}.$$

The gcd of the numerator and the denominator on the right must divide $2^3 \cdot 17$:

$$(-8r + s)(-8r^2 - 2rs + 2s^2) + (18r - 2s)(4r^2 + s^2) = 2^3 \cdot 17r^3 \qquad \text{and}$$
$$(2r + 4s)(-8r^2 - 2rs + 2s^2) + (4r + 9s)(4r^2 + s^2) = 17s^3.$$

If this gcd is even, then s is even and r is odd. If s is not divisible by 4, then $v_2(-8r^2 - 2rs + 2s^2) = 2$ and $v_2(4r^2 + s^2) = 3$; this is impossible because v is odd. This shows that $s = 4s_1$, which gives

$$\frac{-8r^2 - 2rs + 2s^2}{4r^2 + s^2} = \frac{-2r^2 - 2rs_1 + 8s_1^2}{r^2 + 4s_1^2}$$

with odd denominator. Depending on whether the gcd is divisible by 17 or not, we obtain one of the following two systems of equations.

$$\left\{ \begin{array}{l} u_1^2 = -2r^2 - 2rs_1 + 8s_1^2 \\ v^2 = r^2 + 4s_1^2 \end{array} \right\} \qquad \text{or} \qquad \left\{ \begin{array}{l} 17u_1^2 = -2r^2 - 2rs_1 + 8s_1^2 \\ 17v^2 = r^2 + 4s_1^2 \end{array} \right\}.$$

In the case that the gcd is odd, we obtain in a similar way

$$\left\{ \begin{array}{l} u_1^2 = -8r^2 - 2rs + 2s^2 \\ v^2 = 4r^2 + s^2 \end{array} \right\} \qquad \text{or} \qquad \left\{ \begin{array}{l} 17u_1^2 = -8r^2 - 2rs + 2s^2 \\ 17v^2 = 4r^2 + s^2 \end{array} \right\}.$$

In each case u_1 must be even. We write $u_1 = 2u_2$ and divide by 2. The left hand side of the first equation is still even, so we obtain that (depending on the case we are in)

$$r(r + s_1) \equiv 0 \bmod 2 \qquad \text{or} \qquad s(s - r) \equiv 0 \bmod 2.$$

In the first case r is odd, in the second case s is odd. Then s_1 or r, respectively, is also odd. But then $r^2 + 4s_1^2 \equiv 5 \bmod 8$, respectively, $4r^2 + s^2 \equiv 5 \bmod 8$, a contradiction to $v^2 \equiv 17v^2 \equiv 1 \bmod 8$. Hence equation (22.3) has no nontrivial integral solutions. ♣

In principle, one can try to show in general in a similar way that an element of $S_{\hat{\phi}}$ is not in the image of δ. A solution of $w^2 = du^4 + au^2v^2 + cv^4$ also provides a solution of $w^2 = dX^2 + aXY + cY^2$. If this equation has solutions everywhere locally, then it also has a nontrivial integral solution (Hasse-Minkowski Theorem); in the example, one such solution was $(X, Y, w) = (8, 1, 2)$. We can then rationally parameterize the solutions of this equation in w, X, Y (in the example, by the parameter t); this results in one or several systems of equations of the form

$$u^2 = Q_1(r, s), \qquad v^2 = Q_2(r, s)$$

with binary quadratic forms Q_1 and Q_2, which one can then again check for everywhere local solubility. But there are also examples, for which this is not sufficient to determine the rank.

If one has determined successfully the images of δ and of δ', then one also has found for every $d \in \operatorname{im}(\delta)$ a point $P_d \in E(\mathbb{Q})$ with $\delta(P_d) = d$ and for every $d' \in \operatorname{im}(\delta')$ a point $Q_{d'} \in E'(\mathbb{Q})$ with $\delta'(Q_{d'}) = d'$. Then the set

$$R = \{P_d \mid d \in \operatorname{im}(\delta)\} + \{\hat{\phi}(Q_{d'}) \mid d' \in \operatorname{im}(\delta')\}$$

contains a complete system of representatives of the cosets of $2E(\mathbb{Q})$ in $E(\mathbb{Q})$. This follows from the exact sequence (this means that the image of each homomorphism equals the kernel of the next)

$$\frac{E'(\mathbb{Q})}{\phi(E(\mathbb{Q}))} \longrightarrow \frac{E(\mathbb{Q})}{2E(\mathbb{Q})} \longrightarrow \frac{E(\mathbb{Q})}{\hat{\phi}(E'(\mathbb{Q}))} \longrightarrow 0\,,$$

which is behind the proof of Lemma 21.1. The first map does not have to be injective (the kernel is generated by $T' + \phi(E(\mathbb{Q}))$ and can have order 1 or 2; compare the proof of Corollary 21.8); therefore there can be two representatives of the same coset among the $\hat{\phi}(Q_{d'})$. In any case, Theorem 20.2 together with the explicit values for C and c_P from the proof of Theorem 20.5 provides an explicit bound γ such that $E(\mathbb{Q})$ is generated by R together with all points $P \in E(\mathbb{Q})$ such that $h(P) \le \gamma$. In this way one can determine the structure and generators of $E(\mathbb{Q})$. (In practice one uses stronger bounds for the various constants, leading to a smaller bound γ, but the principle remains the same.)

EX Determination of $E(\mathbb{Q})$

22.16. **Example.** We again consider the curves

$$E\colon y^2 = x(x^2 - 15x + 63) \qquad \text{and} \qquad E'\colon y^2 = x(x^2 + 30x - 27)\,.$$

from Example 22.4. We already had shown there that $\operatorname{rk}(E(\mathbb{Q})) = 1$. More precisely, we had seen that $\operatorname{im}(\delta) = \langle 3, 7\rangle$ and $\operatorname{im}(\delta') = \langle -3\rangle$. The explicit solution of the equation for $3 \in \operatorname{im}(\delta)$ gives the point $P_3 = P = (3, 9) \in E(\mathbb{Q})$. A preimage of 7 is $P_7 = T = (0, 0)$. This implies that $P_{21} = P + T = (21, -63)$ is a preimage of $21 = 3 \cdot 7$. A preimage of $-3 \in \operatorname{im}(\delta')$ is given by $T' = (0, 0) \in E'(\mathbb{Q})$, but its image $\hat{\phi}(T') = O$ in E does not give something new. We can therefore take

$$R = \{O, (0, 0), (3, 9), (21, -63)\}$$

as a system of representatives of $E(\mathbb{Q})/2E(\mathbb{Q})$. We obtain the bound

$$\gamma = 15.518$$

for the height of points that (together with R) generate $E(\mathbb{Q})$ from the proofs of Theorems 20.2 and 20.5. We therefore need to find all rational points $P = (\xi, \eta)$ such that numerator and denominator of ξ are bounded by $\lfloor e^\gamma \rfloor = 5\,486\,637$. This looks worse than it is: the program `ratpoints` does that in less than one second and finds 48 rational points of height $\le \gamma$. These points are all contained in the subgroup generated by T and P. This implies that

$$E(\mathbb{Q}) = \langle T, P\rangle \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}\,. \qquad \clubsuit$$

If the coefficients are a bit larger, then the bound γ that we obtain from Theorems 20.2 and 20.5 is no longer practical. There are considerably better bounds for the difference $h(Q) - \hat{h}(Q)$

22.17. **Example.** We continue the study of the curve in the previous example. If P does not generate the free part of $E(\mathbb{Q})$, then we must have $\bar{P} = m\bar{Q}$ with $m \geq 2$ and $Q \in E(\mathbb{Q})$; here we write $\bar{Q}$ for the image of Q in $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}$. This is equivalent to saying that $P = mQ$ or $P+T = mQ$, because $E(\mathbb{Q})_{\text{tors}} = \{O, T\}$, which one can see using the Nagell-Lutz Theorem 18.5. Now P and $P+T$ are both not divisible by 2, since otherwise they would be in the image of $\hat{\phi}$, so δ would map them to 1, which is not the case. So m must be odd. Since then $T = mT$, we must have $P = mQ$ for some Q. Further, $m \geq 3$.

EX Continuation

$E_1 \colon y^2 = x^3 - 12x + 65$ is a short Weierstrass equation for E. The Magma function `SiksekBound` gives the bound

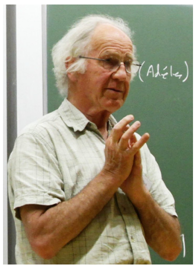
$$\forall Q \in E_1(\mathbb{Q}) \colon h(Q) \leq \hat{h}(Q) + 3.071\,.$$

Then we obtain for the image Q_1 of Q in $E_1(\mathbb{Q})$ (note that the canonical heights of P on E and of the corresponding points on E_1 agree)

$$h(Q_1) \leq \hat{h}(Q_1) + 3.071 = \frac{1}{m^2}\hat{h}(P) + 3.071 \leq \frac{1}{9}\hat{h}(P) + 3.071 \leq 3.082\,.$$

This reduces the bound for the numerator and the denominator of the x-coordinate of Q_1 to 21 and the number of points that need to be considered to 18. ♣

23. The conjecture of Birch and Swinnerton-Dyer



B. Birch
* 1931

As one final topic, I want to explain the conjecture of Birch[12] and Swinnerton-Dyer.

This conjecture essentially says that the numbers $\#\tilde{E}(\mathbb{F}_p)$ of $\mathbb{F}_p$-rational points on the reduction mod p of an elliptic curve E over $\mathbb{Q}$, for all but finitely many p, determine the rank $\operatorname{rk}(E(\mathbb{Q}))$. The heuristic assumption behind this is that when the rank is large, there are "many" rational points on E, whose systematically occurring images under the reduction maps ρ_p should lead to a somewhat larger number of $\mathbb{F}_p$-points than otherwise expected on average. We can express this in a fairly elementary way by writing

$$\#\tilde{E}(\mathbb{F}_p) = p + 1 + A_p$$

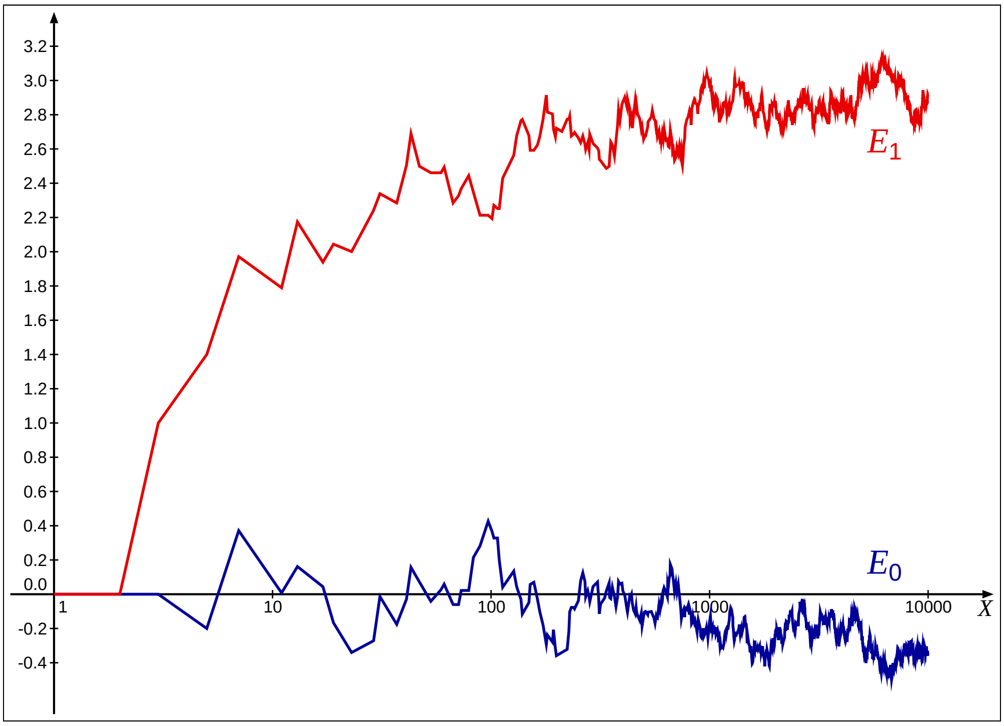
(for primes p at which E has good reduction; see Section 19); then $|A_p| \leq 2\sqrt{p}$ by Hasse's Theorem 13.2 (and $-A_p$ is the trace of Frobenius at p). To capture the deviation from the mean $p + 1$ statistically, we sum A_p/p for all p below a bound X and study the behavior of this sum as $X \to \infty$. For the two curves

H.P.F. Sw.-Dyer
1927 – 2018
Foto © MFO

$$\begin{aligned} E_0 &: y^2 = x^3 - 2x + 1 && \text{with } \operatorname{rk}(E_0(\mathbb{Q})) = 0 \qquad \text{and} \\ E_1 &: y^2 = x^3 - x + 1 && \text{with } \operatorname{rk}(E_1(\mathbb{Q})) = 1, \end{aligned}$$

we find the following behavior of

$$N_E(X) = \sum_{p<X} \frac{A_p}{p}\,.$$



We observe that $N_{E_0}(X)$ remains close to zero, whereas $N_{E_1}(X)$ is clearly growing. We also see that the local behavior of these graphs is quite erratic.

To obtain a "nicer" and more analytic measure of the statistical tendency of the numbers A_p, one instead considers the so-called L-function of E.

23.1. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$, given by a minimal Weierstrass equation (see Definition 19.1). The following infinite product over all prime numbers p converges for $s \in \mathbb{C}$ such that $\operatorname{Re}(s) > \frac{3}{2}$; the thus defined holomorphic function is the (Hasse-Weil-)*L-function of E*. DEF L-function

$$L(E, s) = \prod_p \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{1-2s}} ,$$

where

$$\varepsilon(p) = \begin{cases} 0, & \text{if } p \mid \Delta(E), \\ 1, & \text{otherwise,} \end{cases}$$

and

$$a_p = \begin{cases} -A_p, & \text{if } p \nmid \Delta(E), \\ 0, & \text{if } p \mid c_4(E), c_6(E), \\ \pm 1, & \text{else.} \end{cases}$$

The sign in the last case depends on whether the slopes of the two tangents at the singular point of the reduced curve mod p are both defined over $\mathbb{F}_p$ $(+1)$ or not (-1). ♢

In the case of good reduction (i.e., $p \nmid \Delta(E)$) a_p is the trace of Frobenius, and the denominator of the corresponding factor in the product is the "reciprocal characteristic polynomial of Frobenius" evaluated at p^{-s}. This name comes from the fact that $X^2 - a_p X + p$ is the (reduced) characteristic polynomial of the Frobenius endomorphism ϕ in the endomorphism ring of the mod p reduced elliptic curve $\tilde{E}$ (since $a_p = \phi + \hat{\phi}$ is the trace and $p = \deg(\phi) = \phi\hat{\phi}$ is the norm of ϕ).

In the case of bad reduction one distinguishes between "multiplicative" and "additive" reduction, according to whether the (unique) singular point of the reduced curve is a simple double point (which locally looks like $y^2 = \lambda x^2$ with $\lambda \neq 0$ when $p \neq 2$) or a cusp (which locally looks like $y^2 = x^3$). The usual composition law that defines the group structure still works for singular curves given by a Weierstrass equation, as long as one excludes the singular point. In the case of a simple double point and over an algebraically closed field k, one obtains a group that is isomorphic to the multiplicative group $k^\times$, whereas the group one obtains in the case of a cusp is isomorphic to the additive group of k. This explains the names of the reduction types. In the case of additive reduction, the factor in the product defining the L-function is simply 1. In the case of multiplicative reduction, one distinguishes additionally between *split* and *non-split* multiplicative reduction. In this case, the reduced curve has two tangents at the singular point. If their slopes are defined over $\mathbb{F}_p$, the reduction is split, and the factor in the product is $1/(1-p^{-s})$; the group then is $\mathbb{F}_p^\times$. In the other case, the factor is $1/(1+p^{-s})$, and the group is the subgroup of order $p+1$ of $\mathbb{F}_{p^2}^\times$. In all cases (always assuming that the equation defining E is minimal) we have that $\tilde{E}(\mathbb{F}_p) = 1 + p - a_p$.

The claim that the product converges for $\operatorname{Re}(s) > \frac{3}{2}$ follows from Hasse's Theorem 13.2: $|a_p| \le 2\sqrt{p}$ implies for $\operatorname{Re}(s) > \frac{1}{2}$ that

$$\left| \frac{1}{1 - a_p p^{-s} + \varepsilon(p) p^{1-2s}} - 1 \right| \ll p^{\frac{1}{2} - \operatorname{Re}(s)} ,$$

[12] Picture © W. Stein

and $\sum_p p^{\frac{1}{2}-\mathrm{Re}(s)} < \infty$ as soon as $\mathrm{Re}(s) > \frac{3}{2}$. (An infinite product $\prod_j a_j$ converges (absolutely) if and only if the infinite series $\sum_j (a_j - 1)$ converges (absolutely). This can be seen by taking logarithms since $\log(1+x) \sim x$ for small x.)

If we take $s = 1$ in the product (which is not really allowed!), then formally we obtain

$$\begin{aligned}\log L(E,1) \text{ ``$=$'' } \log \prod_p \frac{1}{1 - \frac{a_p}{p} + \frac{\varepsilon(p)}{p}} \\ = -\sum_p \log\Bigl(1 - \frac{a_p}{p} + \frac{\varepsilon(p)}{p}\Bigr) = -\sum_p \Bigl(\frac{A_p}{p} + O\Bigl(\frac{1}{p}\Bigr)\Bigr),\end{aligned}$$

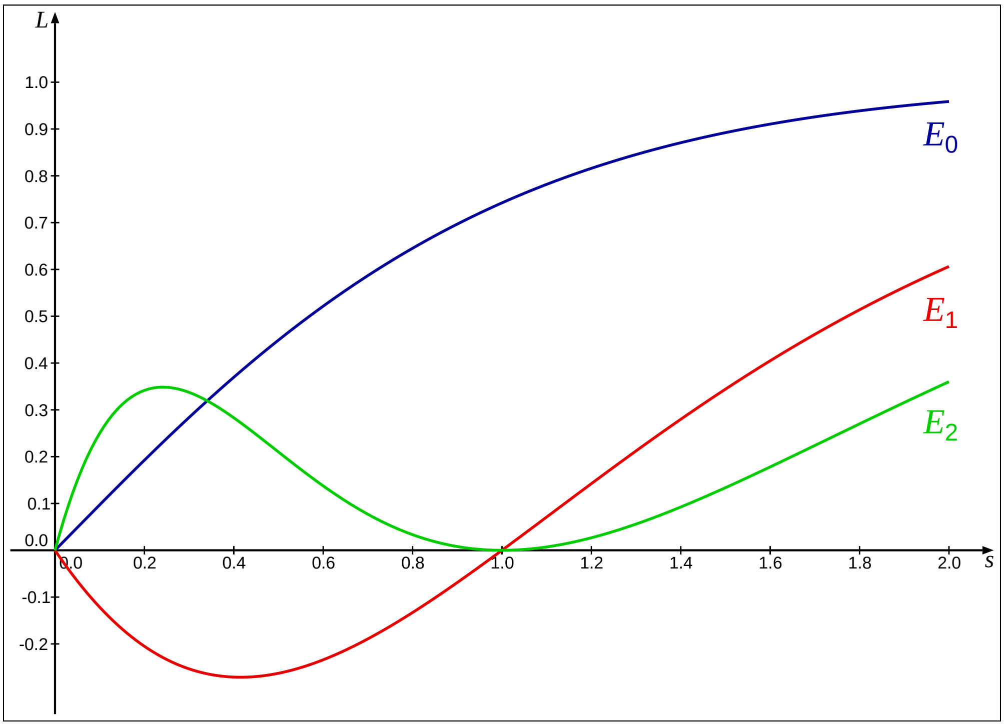
so that we are led to expect that the behavior of $L(E,s)$ close to $s = 1$ should be related to the growth of $N_E(X)$ as $X \to \infty$. The problem with this is that $L(E,s)$ is not even defined at $s = 1$! Still, around 1965, Birch and Swinnerton-Dyer have proposed the following conjecture.

23.2. **Conjecture.** *Let E be an elliptic curve over $\mathbb{Q}$. Then the L-function $L(E,s)$ can be continued to a holomorphic function in a neighborhood of $s = 1$, and*

$$\mathrm{ord}_{s=1} L(E,s) = \mathrm{rk}(E(\mathbb{Q}))\,.$$

CONJ
weak BSD conjecture

This is illustrated for the curves E_0 and E_1 and a further curve E_2 with $\mathrm{rk}(E_2(\mathbb{Q})) = 2$ in the following picture that shows the graphs of the associated L-functions on the positive real axis.



For elliptic curves E with complex multiplication, i.e., such that $\mathrm{End}_{\bar{\mathbb{Q}}}(E)$ is strictly larger than $\mathbb{Z}$ (for examples curves of the form $y^2 = x^3 + ax$ or $y^2 = x^3 + b$), it was already known that their L-function agrees with a certain other L-function (a so-called Hecke L-function), for which Deuring had shown in 1941 that it has a holomorphic continuation to all of $\mathbb{C}$. The conjecture was based on many numerical examples for such curves that had been computed on the EDSAC 2 computer in Cambridge.

The first part (holomorphic continuation) of the conjecture, which in fact already goes back to Hasse, has been proved by now (as a consequence of the proof of the Modularity Conjecture), so that the second and main part now indeed makes sense for all elliptic curves over $\mathbb{Q}$. This proof again is based on showing that the L-function of an elliptic curve can be identified with a different kind of L-function, namely that associated to a modular form of weight 2 for a suitable subgroup of the group $\Gamma = \mathrm{SL}(2, \mathbb{Z})$.

23.3. **Definition.** We write $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ for the upper half-plane. For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and $z \in \mathbb{H}$, we set

$$\gamma z = \frac{az+b}{cz+d}.$$

It is a fact that $\gamma z \in \mathbb{H}$ (exercise). ♢

DEF upper half-plane

23.4. **Definition.** If $f \colon \mathbb{H} \to \mathbb{C}$ satisfies $f(z+m) = f(z)$ for all z and some $m \in \mathbb{Z}_{\geq 1}$, then it has a *Fourier expansion*

$$f(z) = \sum_{n=-\infty}^{\infty} a_n q^{n/m} \qquad \text{with } q^r := e^{2\pi \boldsymbol{i} r z}.$$

This is called the *q-expansion* of f. We say that *f is holomorphic/vanishes at $\boldsymbol{i}\infty$* if $a_n = 0$ for all $n < 0/n \leq 0$. If f is holomorphic at $\boldsymbol{i}\infty$, then we define $f(\boldsymbol{i}\infty) := a_0$. ♢

DEF holomorphic at $\boldsymbol{i}\infty$

23.5. **Definition.** Let $N \in \mathbb{Z}_{\geq 1}$. We define the subgroup $\Gamma_0(N)$ of Γ by

$$\Gamma_0(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma \;\middle|\; c \equiv 0 \bmod N \right\}. \qquad ♢$$

DEF $\Gamma_0(N)$

(There are also subgroups $\Gamma_1(N)$ and $\Gamma(N)$.)

23.6. **Definition.** Let $f \colon \mathbb{H} \to \mathbb{C}$ be holomorphic, let $k \in \mathbb{Z}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Then we define $f|_k\gamma \colon \mathbb{H} \to \mathbb{C}$ by

$$(f|_k\gamma)(z) = (cz+d)^{-k} f(\gamma z);$$

this is again a holomorphic function on $\mathbb{H}$. ♢

DEF $f|_k\gamma$

Using this notation, we can define modular forms of weight k for $\Gamma_0(N)$.

23.7. **Definition.** Let $N \in \mathbb{Z}_{\geq 1}$ and let $k \in \mathbb{Z}$. A *modular form of weight k for $\Gamma_0(N)$* is a holomorphic function $f \colon \mathbb{H} \to \mathbb{C}$ with the following properties.

(1) $f|_k\gamma = f$ for all $\gamma \in \Gamma_0(N)$.

(2) $f|_k\gamma$ is holomorphic at $\boldsymbol{i}\infty$ for all $\gamma \in \Gamma$.

If in addition $f|_k\gamma$ vanishes at $\boldsymbol{i}\infty$ for all $\gamma \in \Gamma$, then f is a *cusp form of weight k for $\Gamma_0(N)$*. ♢

DEF modular form for $\Gamma_0(N)$

Since $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$ for all N and $(f|_kT)(z) = f(z+1)$, modular forms for $\Gamma_0(N)$ are periodic with period 1 and so have a q-expansion of the form

$$f(z) = \sum_{n=0}^{\infty} a_n(f) q^n \qquad \text{with } q = e^{2\pi \boldsymbol{i} z}.$$

If f is a cusp form, so in particular $a_0(f) = 0$, then the Mellin transform of f along the positive imaginary axis defines a Dirichlet series (up to the factor $(2\pi)^{-s}\Gamma(s)$)

$$\int_0^\infty f(\boldsymbol{i}t)t^s \,\frac{dt}{t} = \sum_{n=1}^\infty a_n(f) \int_0^\infty e^{-2\pi nt} t^s \,\frac{dt}{t} \overset{2\pi nt \leftarrow u}{=} (2\pi)^{-s} \sum_{n=1}^\infty \frac{a_n(f)}{n^s} \int_0^\infty e^{-u} u^s \,\frac{du}{u}$$
$$= (2\pi)^{-s}\Gamma(s) \sum_{n=1}^\infty \frac{a_n(f)}{n^s}\,.$$

Here $\Gamma(s)$ is the Gamma function.

Let $\mathcal{S}(k, \Gamma_0(N))$ denote the $\mathbb{C}$-vector space of cusp forms of weight k for $\Gamma_0(N)$. For each integer $n \geq 1$ one can define an endomorphism T_n of $\mathcal{S}(k, \Gamma_0(N))$, a so-called Hecke operator. These endomorphisms commute in pairs; they generate a $\mathbb{C}$-subalgebra of the endomorphism ring $\operatorname{End}_{\mathbb{C}}\bigl(\mathcal{S}(k, \Gamma_0(N))\bigr)$. A cusp form f that is a simultaneous eigenvector of all these Hecke operators is called a *(Hecke) eigenform*. An eigenform f is *normalized* if $a_1(f) = 1$. Then $T_n f = a_n(f)f$. If a normalized eigenform f is not of the form $f(z) = \tilde{f}(dz)$ for some $\tilde{f} \in \mathcal{S}(k, \Gamma_0(M))$ and some $d \in \mathbb{Z}_{\geq 1}$ such that $dM \mid N$ and $M < N$, then f is called a *newform* of level N ("new" since f does not come from a lower, "old", level M). Then we have the following.

DEF
eigenform
normalized
newform

23.8. **Theorem.** *Let $N \in \mathbb{Z}_{\geq 1}$ and let $f \in \mathcal{S}(2, \Gamma_0(N))$ be a newform with coefficients $a_n(f) \in \mathbb{Z}$ for all $n \geq 1$. Then there exists an elliptic curve E over $\mathbb{Q}$ such that*

$$L(f, s) := \sum_{n=1}^\infty \frac{a_n(f)}{n^s} = L(E, s)\,.$$

The function $L(f, s)$ can be holomorphically extended to all of $\mathbb{C}$ and satisfies the functional equation

$$\Lambda(f, 2 - s) = \pm\Lambda(f, s)$$

(for one of the signs), where

$$\Lambda(f, s) := (2\pi)^{-s}\Gamma(s)N^{s/2}L(f, s)\,.$$

THM
ell. curve
from newform

By expanding $(1 - a_p p^{-s} + \varepsilon(p)p^{1-2s})^{-1}$ as a formal power series in p^{-s} and then formally expanding the infinite product from the definition of $L(E, s)$ we also obtain a Dirichlet series.

In particular, $L(E, s)$ (for E as in the theorem above) also has a holomorphic continuation to all of $\mathbb{C}$ and satisfies the functional equation. Since isogenous elliptic curves have the same L-function (see Theorem 13.3), E is here only determined up to isogeny.

The number N (the level of f) then agrees with the *conductor* N_E of E. The conductor has the following properties.

DEF
conductor N_E

(1) N_E divides the minimal discriminant of E.

(2) The prime divisors of N_E are exactly the primes of bad reduction for E.

(3) If E has multiplicative reduction at p, then $v_p(N_E) = 1$.

(4) If E has additive reduction at p, then $v_p(N_E) \geq 2$, with equality when $p \geq 5$. Also, $v_2(N_E) \leq 8$ and $v_3(N_E) \leq 5$; the precise value can determined by an algorithm.

23.9. Example. One can show that the function with the q-expansion

EX
ell. curve
from newform

$$\begin{aligned} f(z) &= q\prod_{n=1}^{\infty}\big((1-q^n)(1-q^{11n})\big)^2 \\ &= q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 - 2q^9 - 2q^{10} + q^{11} - 2q^{12} + 4q^{13} + \dots \end{aligned}$$

is an element of $\mathcal{S}(2, \Gamma_0(11))$ and that this space has dimension 1. Since 11 is prime, f must then be a newform. There is an isogeny class of elliptic curves E with $N_E = 11$; one of these curves is

$$E\colon y^2 + y = x^3 - x^2 .$$

We indeed have that

$$\begin{aligned} \#\tilde{E}(\mathbb{F}_2) &= 5 = 2 + 1 - (-2) \\ \#\tilde{E}(\mathbb{F}_3) &= 5 = 3 + 1 - (-1) \\ \#\tilde{E}(\mathbb{F}_5) &= 5 = 5 + 1 - 1 \\ \#\tilde{E}(\mathbb{F}_7) &= 10 = 7 + 1 - (-2) \\ \#\tilde{E}(\mathbb{F}_{13}) &= 10 = 13 + 1 - 4 \end{aligned}$$

etc. (The numbers of points are all divisible by 5 since $\#E(\mathbb{Q})_{\text{tors}} = 5$.) We therefore obtain something like an explicit formula for these numbers! ♣

23.10. Definition. An elliptic curve E over $\mathbb{Q}$ for which there exists a newform f such that $L(f, s) = L(E, s)$ is *modular.* ♢

DEF
modular
ell. curve

So for modular elliptic curves E the expression "$\operatorname{ord}_{s=1} L(E, s)$" in Conjecture 23.2 makes sense.

There was a conjecture (formulated in 1958 by Taniyama and Shimura) that at the time appeared to be quite far-fetched, which claims that *all* elliptic curves over $\mathbb{Q}$ should be modular. This conjecture was finally proved in the mid-1990s first by Wiles and Taylor for "semistable" elliptic curves (these are curves that have either good or multiplicative reduction at all primes p) and then in 2001 by Breuil, Conrad, Diamond and Taylor[13] for all elliptic curves over $\mathbb{Q}$:

23.11. Theorem. *Let E be an elliptic curve over $\mathbb{Q}$. Then E is modular.*

THM
Modularity
Theorem

Wiles's motivation for proving modularity of semistable elliptic curves over $\mathbb{Q}$ came from the fact that it was known that this would imply Fermat's Last Theorem.

23.12. Corollary. *The equation*

COR
Fermat's Last
Theorem

$$x^n + y^n = z^n$$

has no solution in integers $x, y, z \neq 0$ and $n \geq 3$.

It is sufficient to consider $n = 4$ or $n = p$ an odd prime. The case $n = 4$ was already dealt with by Fermat who showed the stronger statement that the equation

$$w^2 = u^4 + v^4 \tag{23.1}$$

[13] C. Breuil, B. Conrad, F. Diamond, R. Taylor: *On the modularity of elliptic curves over $\mathbb{Q}$: Wild 3-adic exercises*, J. Amer. Math. Soc. **14** (2001), 843–939.

has no nontrivial (meaning that $u, v, w \neq 0$) integral solution. This equation is of the form that we saw in the context of computing Selmer groups. The corresponding elliptic curve is

$$E_4 \colon y^2 = x^3 + x\,,$$

for which we have shown in Example 22.1 that $E_4(\mathbb{Q}) = \{O, (0,0)\}$. Every nontrivial solution (u, v, w) of (23.1) yields a rational point on E with x-coordinate $(u/v)^2 \neq 0, \infty$; however, such points do not exist.

The case $n = 3$ might also have been dealt with already by Fermat; in any case, Euler gave a proof. The projective plane curve $X^3 + Y^3 = Z^3$ is isomorphic as a curve to the elliptic curve

$$E_3 \colon y^2 = x^3 - 432\,,$$

for which one can also show that it has rank 0. Knowing this, one easily finds that $E_3(\mathbb{Q}) = \{O, (12, 36), (12, -36)\}$, which shows that the original curve also has only the three rational points $(1 : -1 : 0)$, $(1 : 0 : 1)$ and $(0 : 1 : 1)$.

We can also deal with $n = 7$ using what we have learnt in this course.

EX Fermat for $n = 7$

23.13. **Example.** Changing the sign of z, we can write the Fermat equation for $n = 7$ more symmetrically as $x^7 + y^7 + z^7 = 0$. This equation defines a smooth plane projective curve C_7 (of genus $6 \cdot 5/2 = 15$), which has rational points such that $xyz \neq 0$ if and only if Fermat's equation has nontrivial integral solutions.

There is a natural action of the symmetric group S_3 on C_7, by permuting the coordinates. "Dividing out" this action results in a morphism $\psi \colon C_7 \to E_7 = C_7/S_3$, where

$$E_7 \colon y^2 = x(x^2 - 3 \cdot 7^2 x + 2^4 \cdot 7^3)$$

is an elliptic curve. This curve is of the kind we have studied in the context of determining the group of rational points. We have $a = -3 \cdot 7^2$ and $b = 2^4 \cdot 7^3$, so $a^2 - 4b = -7^3$. We find that the Selmer groups are

$$S_{\hat{\phi}} = \langle 7 \rangle \qquad \text{and} \qquad S_{\phi} = \langle -7 \rangle$$

(generated by the images of the points of order 2). This shows that $\operatorname{rk}(E_7(\mathbb{Q})) = 0$, and then it is easy to determine

$$E_7(\mathbb{Q}) = \{O, (0,0)\}\,.$$

So, if $P = (\xi : \eta : \zeta)$ is a rational point on C_7, its image $\psi(P)$ (as a point on $E_7 \subset \mathbb{P}^2$) must be either $O = (0 : 1 : 0)$ or $(0,0) = (0 : 0 : 1)$; in any case, the first coordinate must be zero. This first coordinate is given by

$$(x + y)(x + z)(y + z)(x + y + z)(xy + xz + yz)$$

(at least in one representation of ψ). So one of these factors must vanish, and one sees easily that this is only possible for a rational point when one of the coordinates is zero. ♣

Using Fermat's and Euler's results, we can assume that $n = p \geq 5$ is prime. Now the idea (going back to Frey[14]) is to associate to a putative integral solution

$$a^p + b^p = c^p$$

[14] G. Frey: *Links between stable elliptic curves and certain diophantine equations*, Annales Universitatis Saraviensis. Series Mathematicae **1** (1986), 1–40.

with $a, b, c \neq 0$ and, without loss of generality, $\gcd(a, b, c) = 1$ the elliptic curve ("Frey curve")

$$E_{a,b,c} \colon y^2 = x(x + a^p)(x - b^p)\,,$$

whose discriminant is $\Delta(E_{a,b,c}) = -16(abc)^{2p}$. Possibly after permuting $(a, b, -c)$ and/or simultaneously changing all signs, the equation of $E_{a,b,c}$ is a minimal Weierstrass equation, which has bad multiplicative reduction at 2 and at all odd primes dividing abc (and good reduction at all other primes, so it is semistable). One now considers the action of the absolute Galois group $G_{\mathbb{Q}}$ of $\mathbb{Q}$ on the p-torsion $E_{a,b,c}[p]$. The facts that $E_{a,b,c}$ is modular and semistable, and $v_q(\Delta(E_{a,b,c})) \in p\mathbb{Z}$ for all primes $q \geq 3$ then imply by a result due to Ribet that there would have to exist a newform of level 2 and weight 2 whose q-expansion coefficients are congruent mod p to those of the newform belonging to $E_{a,b,c}$. However, $\mathcal{S}(2, \Gamma_0(2)) = \{0\}$, so there do not even exist newforms of level 2 and weight 2. This contradiction shows that the solution, which was the starting point of our considerations, cannot in fact exist.

The Modularity Theorem 23.11 also implies that we can find *all* elliptic curves over $\mathbb{Q}$ that have conductor N by first determining the newforms of level N with integral coefficients and then finding the by Theorem 23.8 associated elliptic curves (for both steps, algorithms exist; see for example [Cre]). In this way, a database of all elliptic curves E over $\mathbb{Q}$ with $N_E \leq 500\,000$ (plus some further curves) was constructed.

Before we get to what is known on the Birch–Swinnerton-Dyer Conjecture, we want to state its "strong" version. This involves a number of objects that we need to introduce first.

23.14. **Definition.** Let $E \colon y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ be a minimal Weierstrass equation of an elliptic curve over $\mathbb{Q}$. The *real period* $\Omega(E)$ of E is the integral

DEF real period

$$\Omega(E) = \int\limits_{E(\mathbb{R})} \left|\frac{dx}{2y + a_1x + a_3}\right|. \qquad \diamondsuit$$

The differential form

$$\omega_E = \frac{dx}{2y + a_1x + a_3} = \frac{dy}{3x^2 + 2a_2x + a_4 - a_1y}$$

is also called the *invariant differential* of E. It has the property that it remains invariant under addition of an arbitrary point of E: $\tau_P^*\omega_E = \omega_E$, where $P \in E$ and $\tau_P \colon E \to E$, $Q \mapsto P + Q$.

DEF invariant differential

23.15. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$. Let $r = \operatorname{rk}(E(\mathbb{Q}))$ and let $P_1, \ldots, P_r \in E(\mathbb{Q})$ be such that their images in $E(\mathbb{Q})/E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}^r$ form a basis. Then

DEF regulator

$$R(E) = \det\bigl(\langle P_i, P_j\rangle\bigr)_{1 \leq i,j \leq r}$$

is the *regulator* of E. Here

$$\langle P, Q\rangle = \tfrac{1}{2}\bigl(\hat{h}(P + Q) - \hat{h}(P) - \hat{h}(Q)\bigr)$$

is the symmetric bilinear form associated to the quadratic form $\hat{h}$. The factor $\frac{1}{2}$ is there so that we have $\langle P, P\rangle = \hat{h}(P)$. $\diamondsuit$

The regulator does not depend on the choice of basis. To see that, let G be the (Gram) matrix in the definition and let $T \in \mathrm{GL}(r, \mathbb{Z})$ be the change-of-basis matrix. The determinant of G changes upon changing the basis into

$$\det(T^\top G T) = \det(T)^2 \det(G) = (\pm 1)^2 \det(G) = \det(G)\,.$$

For the following we need the field $\mathbb{Q}_p$ of p-adic numbers. In a similar way as $\mathbb{R}$, it is a completion of $\mathbb{Q}$, but instead of the usual absolute value, one uses the p-adic absolute value

$$|x|_p = \begin{cases} 0, & \text{if } x = 0, \\ p^{-v_p(x)}, & \text{else} \end{cases}$$

to define the metric ($d(x, y) = |x - y|_p$), with respect to which one constructs the completion. See for example Section 7 in the (German) course notes "Diophantische Gleichungen".

23.16. Definition. Let E be an elliptic curve over $\mathbb{Q}$ given by a minimal Weierstrass equation. Let p be a prime number and let $\tilde{E}$ be the (not necessarily elliptic) curve over $\mathbb{F}_p$ defined by the mod p reduced equation. We consider E as an elliptic curve over the field $\mathbb{Q}_p$ of p-adic numbers; then we have the reduction map $\rho \colon E(\mathbb{Q}_p) \to \tilde{E}(\mathbb{F}_p)$. We define

DEF
Tamagawa number

$$E^{(0)}(\mathbb{Q}_p) = \{P \in E(\mathbb{Q}_p) \mid \rho(P) \in \tilde{E}(\mathbb{F}_p) \text{ is non-singular}\}\,;$$

then $E^{(0)}(\mathbb{Q}_p)$ is a subgroup of finite index in $E(\mathbb{Q}_p)$. The *Tamagawa number* of E at p then is the index

$$c_p(E) = \bigl(E(\mathbb{Q}_p) : E^{(0)}(\mathbb{Q}_p)\bigr)\,. \qquad \diamondsuit$$

If E has good reduction at p, then $E^{(0)}(\mathbb{Q}_p) = E(\mathbb{Q}_p)$ and so $c_p(E) = 1$. Therefore only finitely many of the $c_p(E)$ are not equal to 1, hence the product over all primes (the *Tamagawa product* of E)

$$c(E) = \prod_p c_p(E)$$

makes sense.

The Tamagawa numbers have the following properties.

(1) If E has good reduction at p, then $c_p(E) = 1$.

(2) If E has bad reduction that is not split multiplicative at p, then $c_p(E) \le 4$.

(3) If E has split multiplicative reduction at p, then $c_p(E) = v_p(\Delta(E)) = -v_p(j(E))$.

(4) For given E and p, $c_p(E)$ can be explicitly computed.

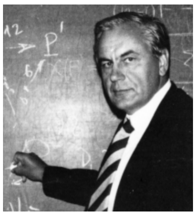
The last object that we need is the *Shafarevich-Tate group* (or also *Tate-Shafarevich group*) of E.

23.17. Definition. Let E be an elliptic curve over $\mathbb{Q}$. A *torsor* under E is a (smooth) projective (not necessarily plane) curve X defined over $\mathbb{Q}$, together with a morphism $\mu \colon E \times X \to X$ that is also defined over $\mathbb{Q}$ and gives rise to an action of E on X:

DEF
torsor

$$\forall P, Q \in E, x \in X \colon \mu(O, x) = x \quad \text{and} \quad \mu(P + Q, x) = \mu(P, \mu(Q, x))$$

such that the induced action of $E(\bar{\mathbb{Q}})$ on $X(\bar{\mathbb{Q}})$ is transitive with trivial stabilizers (for all $x, y \in X(\bar{\mathbb{Q}})$ there is *exactly one* point $P \in E(\bar{\mathbb{Q}})$ such that $\mu(P, x) = y$).


I.R. Shafarevich
1923 – 2017
Foto © MFO

Two torsors (X,μ) and (X',μ') under E are *isomorphic*, if there exists an isomorphism of curves $\phi\colon X \to X'$ that is defined over $\mathbb{Q}$ and such that the following diagram commutes.

$$\begin{tikzcd} E \times X \arrow[r, "\mathrm{id}\times\phi"] \arrow[d, "\mu"] & E \times X' \arrow[d, "\mu'"] \\ X \arrow[r, "\phi"] & X' \end{tikzcd}$$

A torsor (X,μ) is *trivial*, if $X(\mathbb{Q}) \neq \emptyset$. It is *locally trivial*, if $X(\mathbb{R}) \neq \emptyset$ and $X(\mathbb{Q}_p) \neq \emptyset$ for all primes p. ◊

The trivial torsor is $(E,+)$ (where $+$ is the addition map $E \times E \to E$). A torsor is trivial if and only if it is isomorphic to $(E,+)$. ("$\Rightarrow$": Let $x \in X(\mathbb{Q})$. Then $\phi\colon E \to X$, $P \mapsto \mu(P,x)$, is the desired isomorphism. "$\Leftarrow$": Let $\phi\colon E \to X$ be the isomorphism. Then $x = \phi(O) \in X(\mathbb{Q})$.) Since every curve X always has $\bar{\mathbb{Q}}$-rational points, one also can define a torsor under E to be a pair (X,μ) that is defined over $\mathbb{Q}$ and becomes isomorphic to $(E,+)$ over $\bar{\mathbb{Q}}$.

From two torsors (X,μ) and (X',μ') one can construct a third one, the Baer sum of the two. It is defined as the quotient of $X \times X'$ by the action of E given by

$$P * (x,x') = \bigl(\mu(P,x), \mu'(-P,x')\bigr).$$

The Baer sum is commutative and associative, compatible with isomorphisms and has the trivial torsor as a neutral element (up to isomorphism). Also, every torsor (X,μ) has an inverse up to isomorphism, which is given by $\bigl(X, (P,x) \mapsto \mu(-P,x)\bigr)$. Therefore the set of isomorphism classes of torsors under E forms an abelian group, the Weil-Châtelet group of E.

23.18. **Definition.** Let E be an elliptic curve over $\mathbb{Q}$. The *Shafarevich-Tate group* $Ш(E)$ of E is the subgroup of the Weil-Châtelet group of E consisting of isomorphism classes of locally trivially torsors. ◊

DEF
Shafarevich-Tate group

23.19. **Example.** We have already met some torsors. Let $\hat{\phi}\colon E' \to E$ be an isogeny of degree 2, where $E\colon y^2 = x(x^2+ax+b)$. For $d \in \mathbb{Q}^\times$ we had considered the equation

EX
torsors

$$w^2 = du^4 + au^2v^2 + \frac{b}{d}v^4.$$

This equation defines a smooth curve X_d in a weighted projective plane (we treat w as having degree 2). This curve X_d is a torsor under E'. To see this, one shows that X_d is isomorphic to E' over $\mathbb{Q}(\sqrt{d})$ (since X_d is isomorphic to X_1 over $\mathbb{Q}(\sqrt{d})$, it suffices to show that X_1 is isomorphic to E' over $\mathbb{Q}$). One can choose this isomorphism $\varphi\colon X_d \to E'$ in such a way that it is compatible with $\hat{\phi}$ and the map $X_d \to E$. One then defines μ such that (X_d,μ) is isomorphic to $(E',+)$ over $\mathbb{Q}(\sqrt{d})$:

$$\mu(P,x) = \varphi^{-1}(P + \varphi(x));$$

it remains to show that μ is defined over $\mathbb{Q}$.

The isomorphism class of (X_d,μ) is an element of $Ш(E')$ if and only if X_d is everywhere locally soluble, i.e., if and only if $d \in S_{\hat{\phi}}$. The torsor (X_d,μ) is trivial if and only if X_d has a rational point, i.e., if and only if $d \in \mathrm{im}(\delta)$. This gives an exact sequence

$$E'(\mathbb{Q}) \xrightarrow{\hat{\phi}} E(\mathbb{Q}) \xrightarrow{\delta} S_{\hat{\phi}} \longrightarrow Ш(E').$$

The isogeny $\hat{\phi}$ induces a homomorphism $\hat{\phi}_* : \text{Ш}(E') \to \text{Ш}(E)$. The image of the last map in the sequence above then is exactly the kernel $\text{Ш}(E')[\hat{\phi}_*]$ of $\hat{\phi}_*$. We obtain the equivalence

$$\operatorname{im}(\delta) = S_{\hat{\phi}} \iff \text{Ш}(E')[\hat{\phi}_*] = \{0\}\,.$$

The analogous statement holds for arbitrary isogenies and the associated Selmer groups. ♣

Now we have everything we need to state the strong version of the Birch–Swinnerton-Dyer Conjecture.

23.20. **Conjecture.** *Let E be an elliptic curve over $\mathbb{Q}$ given by a minimal Weierstrass equation. Let $r = \operatorname{rk}(E(\mathbb{Q}))$.*

CONJ strong BSD Conjecture

Then the group $\text{Ш}(E)$ is finite, we have

$$\operatorname{ord}_{s=1} L(E,s) = r\,,$$

and

$$\frac{1}{r!}\Big(\frac{d}{ds}\Big)^r L(E,s)\Big|_{s=1} = \lim_{s\to 1}\frac{L(E,s)}{(s-1)^r} = \Omega(E)c(E)\,\frac{R(E)\#\text{Ш}(E)}{(\#E(\mathbb{Q})_{\text{tors}})^2}\,. \tag{23.2}$$

What is known about this? First of all, the conjecture holds for an elliptic curve E if and only if it holds for an isogenous curve E'. In this case we have $L(E,s) = L(E',s)$; the claim here therefore is that the expressions on the right hand side of Equation (23.2) for E and E' agree (it is easy to see that $\operatorname{rk}(E(\mathbb{Q})) = \operatorname{rk}(E'(\mathbb{Q}))$). This is by no means obvious, as each single term can change!

The most important result in direction of the conjecture was obtained in 1988 by Kolyvagin:

23.21. **Theorem.** *Let E be a (modular) elliptic curve over $\mathbb{Q}$ such that*

THM BSD for ord ≤ 1

$$\operatorname{ord}_{s=1} L(E,s) \le 1\,.$$

Then

$$\operatorname{ord}_{s=1} L(E,s) = \operatorname{rk}(E(\mathbb{Q}))\,,$$

$\text{Ш}(E)$ is finite, and (23.2) holds up to a nonzero rational factor whose numerator and denominator are divisible only by primes from a finite set that can be explicitly determined.

M. Bhargava
* 1974

Kolyvagin had proved this result under the assumption that E is modular; this was some time before the Modularity Theorem 23.11 was established.

It is expected that the condition $\operatorname{ord}_{s=1} L(E,s) \le 1$ holds for "almost all" elliptic curves E. More precisely, the conjecture is that the proportion of elliptic curves with conductor $\le X$ that satisfy this tends to 1 as $X \to \infty$. The best result in this direction is due to Bhargava and various coauthors and says that the proportion of such curves is larger than 0.66 when X is sufficiently large. In this sense, Theorem 23.21 gets fairly close to a complete proof. Based on this result and many further improvements regarding the prime numbers that can occur in the "error factor", by work of many mathematicians in a number of papers the following could be established.

23.22. **Theorem.** *Let E be an elliptic curve over $\mathbb{Q}$ such that*

$$N_E \le 5000 \qquad \textit{and} \qquad \operatorname{rk}(E(\mathbb{Q})) \le 1$$

(This excludes only 691 of the in total 31 073 curves E with $N_E \le 5000$). Then the strong BSD Conjecture holds for E.

THM BSD for $N_E \le 5000$

On the other hand, not much is known when $\operatorname{rk}(E(\mathbb{Q})) \ge 2$ or $\operatorname{ord}_{s=1} L(E,s) \ge 2$.

(1) One can determine the sign in the functional equation of $L(E,s)$ for any given E. When it is $+1$, then $\operatorname{ord}_{s=1} L(E,s)$ is even, otherwise odd.

(2) One can decide whether $L(E,1) = 0$ (in the even case), resp., $L'(E,1) = 0$ (in the odd case) or not. One can also verify numerically that $L^{(n)}(E,1) \ne 0$. In this way it is possible to determine $\operatorname{ord}_{s=1} L(E,s)$ if this order is at most 3. On the other hand, there is so far no method available to prove that the order is a given number ≥ 4. In particular, it is not currently possible to verify the weak BSD Conjecture when $\operatorname{rk} E(\mathbb{Q}) \ge 4$.

(3) Not a single elliptic curve E over $\mathbb{Q}$ with $\operatorname{rk}(E(\mathbb{Q})) \ge 2$ is known for which it could be shown that $Ш(E)$ is finite.

(4) There are no candidates for potential counterexamples to the BSD Conjecture.

The Clay Foundation offers a prize of one million US dollars for a proof of the weak BSD Conjecture for all elliptic curves over $\mathbb{Q}$ (there is a more general version for abelian varieties over algebraic number fields). This is one of the Clay Millennium Problems, along with, for example, the Riemann Hypothesis on the nontrivial zeros of the Riemann zeta function.

References

[Cas] J.W.S. Cassels: *Lectures on elliptic curves*, London Mathematical Society Student Texts **24**, Cambridge University Press (1991).

[Co1] H. Cohen: *A course in computational algebraic number theory*, Springer GTM **138** (1993).
Online access (from the university network)

[Co2] H. Cohen: *Number Theory. Volume I: Tools and diophantine equations*, Springer GTM **239** (2007).
Online access (from the university network)

[Cre] J.E. Cremona: *Algorithms for modular elliptic curves*, second edition. Cambridge University Press, Cambridge, 1997.
Free online version

[CF+] J.E. Cremona, T.A. Fisher, C. O'Neil, D. Simon und M. Stoll: *Explicit n-descent on elliptic curves. I. Algebra*, J. reine angew. Math. **615**, 121–155 (2008).

[Hus] D. Husem"oller: *Elliptic curves*, Springer GTM **111** (1987).
Online access (from the university network)

[Jae] K. J"anich: *Einf"uhrung in die Funktionentheorie*, Springer Hochschultext (2. Auflage 1980).
Online access (from the university network)

[Kna] A.W. Knapp: *Elliptic curves*, Mathematical Notes **40**, Princeton University Press (1992).

[Si1] J.H. Silverman: *The arithmetic of elliptic curves*, Springer GTM **106** (1986).
Online access (from the university network)

[Si2] J.H. Silverman: *Advanced topics in the arithmetic of elliptic curves*, Springer GTM **151** (1994).
Online access (from the university network)

[ST] J.H. Silverman und J.T. Tate: *Rational points on elliptic curves*, second edition. Springer Undergraduate Texts in Mathematics (2015).
Online access (from the university network)