Cubic curves over finite fields

P. Edera and F. Hoffmann

31st October 2023

1 Rational points over finite fields

In this Section we explore cubic equations within the confines of a finite field, specifically focusing on the field $\mathbb{F}_p$, which corresponds to the integers modulo p. Our exploration will closely align with the content covered in Chapter 4 of the work authored by Silverman and Tate [1]. In general Cubic curves are a special case of elliptic curves when they have a specified point at infinity. Instead, over finite fields, they are a fundamental topic in algebraic geometry and number theory. These curves are a special class of algebraic curves defined by cubic equations over finite fields. In the context of these fields, cubic curves have important applications in cryptography, error-correcting codes, and coding theory.

Consider a prime number p and a polynomial $F(x, y) \in \mathbb{F}_p[x, y]$ with coefficients in $\mathbb{F}_p$, defining a curve

$$C : F(x, y) = 0.$$

Since we cannot visualize things now, we look for solutions (x, y) of this equation with $(x, y) \in \mathbb{F}_{p^n}$, an extension field of $\mathbb{F}_p$ with p^n elements.

Definition 1.1. *A point (x, y) with $x, y \in \mathbb{F}_p$ and $F(x, y) = 0$ is called a rational point of the curve C. We denote the set of all rational points of C by $C(\mathbb{F}_p)$.*

Example 1.2. *Consider*

$$y^2 = x^3 + x + 1$$

over $\mathbb{F}_5$. Now we can simply consider each of the five potential values for x, substitute them into the polynomial $x^3 + x + 1$, and verify whether the outcome is a square within the field $\mathbb{F}_5$. We therefore find nine points, including $\mathcal{O}$ at infinity:

$$C(\mathbb{F}_5) = \{\mathcal{O}, (0, \pm1), (2, \pm1), (3, \pm1), (4, \pm2)\}.$$

When dealing with a non-singular cubic curve C, it becomes possible to establish an addition law on the curve. The rational points on C, in combination with the point $\mathcal{O}$ at infinity, collectively constitute an abelian group under this defined operation. Utilizing images is unnecessary because the procedures and formulas that we have seen for cubic curves over $\mathbb{C}$ also work over $\mathbb{F}_p$. As the field $\mathbb{F}_p$ is of finite size, there exists only a finite number of points that can exist on the curve C. This observation leads to the conclusion that $C(\mathbb{F}_p)$ forms a finite group.

One of the central problems in the study of cubic curves over finite fields is determining the number of rational points on the curve. Therefore we want to find an exact formula or an estimate for the number of points in $C(\mathbb{F}_p)$. First and foremost, it's important to emphasize that C invariably includes the point $\mathcal{O}$ at infinity. Concerning the remaining points, we can assert that they adhere to the pattern (x, y), where x and y are elements of the finite field $\mathbb{F}_p$. Since the Cartesian product $\mathbb{F}_p \times \mathbb{F}_p$ is a finite set, our task is simply to examine whether the function $F(x, y)$ equals zero for all potential combinations of x and y drawn from $\mathbb{F}_p$. An important result that we can use to achieve our goal is the Hesse-Weil theorem. This theorem provides us an upper bound on the number of rational points, which is related to the size of the finite field.

2 The Hasse-Weil theorem

We now consider the curve C given by

$$C : y^2 = f(x),$$

where $f(x)$ is a polynomial with coefficients in $\mathbb{F}_p$, and we try to estimate the number of rational points of the curve C. To do this, we first assume that $p \neq 2$. It is now important to define a new concept:

Definition 2.1. *An integer x is called a quadratic residue modulo p if there exists an integer a such that:*

$$a^2 = x \ (mod \, p).$$

Otherwise, x is called a quadratic nonresidue modulo p.

The number of quadratic residues is related to Euler's criterion, which determines whether an element in a finite field is a quadratic residue modulo the field's characteristic prime. In other words, it helps determine if an integer can be expressed as a perfect square modulo a prime. It turns out that the number of quadratic residues in a finite field $\mathbb{F}_{p^n}$ depends on the field's size p^n and its prime characteristic p. In most cases, there are exactly $(p-1)/2$ quadratic residues among the nonzero elements of the field, but this number may vary when the field has characteristic 2 (in this case every non-zero element is a quadratic residue since $a^2 = (-a)^2$ for all elements a in a field of characteristic 2); this is why we have assumed $p \neq 2$. It's important to note that discussions about quadratic residues typically exclude the field's zero element since $0^2 = 0$ for any element in any field. The following proposition summarises what has been said:

Proposition 2.2. *Let $p \geq 3$ be a prime number. Then there are exactly $\frac{p-1}{2}$ non-zero elements of the field $\mathbb{F}_p$ which are squares (i.e. elements of the form x^2 for $x \in \mathbb{F}_p^*$).*

Therefore we can now consider $x \in \mathbb{F}_p$ and think of substituting the different values $x = 0, \ldots, p-1$ into the equation $y^2 = f(x)$. If $f(x)$ equals zero, the only feasible value for y is $y = 0$. In cases where $f(x)$ is non-zero, we can distinguish between two scenarios: either $f(x)$ is one of the $p-1$ residues in $\mathbb{F}_p$, leading to two potential values for y, or $f(x)$ is a nonresidue in $\mathbb{F}_p$, resulting in no solutions for y. Hence, assuming that the values of $f(x)$ are uniformly distributed across the elements of $\mathbb{F}_p$, we can estimate that, on average, each $x \in \mathbb{F}_p$ has approximately one solution. When including the point $\mathcal{O}$ at infinity, this yields a total of $p+1$ points. Importantly, for a separable polynomial $f \in \mathbb{F}_p[x]$, there is no discernible bias in the distribution of f values as squares or non-squares. Consequently, we can estimate that the overall number of rational points on the curve C, including the point at infinity, is approximately $p+1$. This estimation can be made precise using the Hasse-Weil theorem, as discussed in Silverman and Tate [1] on page 120.

Theorem 2.3 (Hasse-Weil)**.** *If C is a non-singular irreducible curve of genus g defined over a finite field $\mathbb{F}_p$, then the number of points on C with coordinates in $\mathbb{F}_p$ is equal to $p + 1 - \epsilon$, where ϵ satisfies $|\epsilon| \leq 2g\sqrt{p}$.*

The full explanation of the genus concept is beyond the scope of this presentation. For now, we will simply mention that every curve of the form $F(x, y) = 0$ is linked to a non-negative number g, referred to as its genus (in a nutshell it is a topological invariant that measures the number of "holes" in the curve's surface). In particular, any non-singular curve given by a cubic equation is a curve of genus 1. Therefore for an elliptic curve C over $\mathbb{F}_p$, we have the estimate

$$\Big||C(\mathbb{F}_p)| - (p+1)\Big| \leq 2\sqrt{p}.$$

3 A theorem of Gauss

We can now give some special cases of Theorem 2.3. Gauss successfully established proofs for specific instances of the theorem. In this section, we delve into one such instance, focusing on the cubic Fermat curve

$$x^3 + y^3 = 1.$$

We consider the curve in its homogeneous form

$$x^3 + y^3 + z^3 = 0$$

and its solutions in the projective sense. Hence we do not count the trivial solution $(0, 0, 0)$ and we identify a solution (x, y, z) with all of its non-zero multiples (ax, ay, az).

Before continuing, we state some definitions and propositions (without proving them because they are not important with the aim of this talk) that will make our life easier for the computations in the proof of Theorem 3.5.

Definition 3.1. *Let $X, Y, Z \subset \mathbb{F}_p$ be subsets of $\mathbb{F}_p$. Then we denote by $[XYZ]$ the number of triples $(x, y, z) \in X \times Y \times Z$ with $x + y + z = 0$.*

Proposition 3.2. *Let $X, Y, Z, W \subset \mathbb{F}_p$. Then it holds:*

(i) If $Z \cap W = \emptyset$, then $[XY(Z \cup W)] = [XYZ] + [XYW]$.

(ii) $\forall a \in \mathbb{F}_p^* : [XYZ] = [aX, aY, aZ]$.

(iii) $[XYZ] = [XZY] = [YXZ] = [YZX] = [ZXY] = [ZYX]$.

Proposition 3.3. *Every line in the projective plane over the finite field $\mathbb{F}_p$ has exactly $p + 1$ points.*

Proposition 3.4. *The multiplicative group $\mathbb{F}_p^*$ of $\mathbb{F}_p$ consisting of the non-zero elements $1, 2, \ldots, p - 1$ with the group operation being multiplication is a cyclic group of order $p - 1$. For instance, $\mathbb{F}_p^*$ has $p - 1$ non-zero elements that can be expressed as powers of a single element.*

Adhering to these established conventions, we are now prepared to present Gauss's theorem; we can find it in Silverman and Tate [1] on page 121.

Theorem 3.5 (Gauss)**.** *Let M_p be the number of projective solutions to the equation*

$$x^3 + y^3 + z^3 = 0$$

with x, y, z in the finite field $\mathbb{F}_p$.

(a) If $p \not\equiv 1$ (mod 3), then $M_p = p + 1$.

(b) If $p \equiv 1$ (mod 3), then there exist integers A and B such that

$$4p = A^2 + 27B^2.$$

The numbers A and B are unique up to changing their signs, and if we fix the sign of A so that $A \equiv 1$ (mod 3), then

$$M_p = p + 1 + A.$$

Remark 3.6. *If $p \equiv 1$ (mod 3) then $4p \equiv 4 \equiv 1$ (mod 3) and also $27B^2 \equiv 0$ (mod 3). Therefore the equation $4p = A^2 + 27B^2$ implies that $A^2 \equiv 1$ (mod 3), which gives us $A \equiv 1$ (mod 3). We can also observe that $A^2 = 4p - 27B^2 < 4p$, and thus $|A| < 2\sqrt{p}$ and so Gauss's theorem is indeed a special case of Theorem 2.3.*

Proof of Theorem 3.5. (a) Let $p \not\equiv 1$ (mod 3). Then we get that 3 does not divide $p - 1 = |\mathbb{F}_p^*|$ and since $\mathbb{F}_p^*$ is a cyclic group by Proposition 3.4, it follows that the map

$$\Phi : \mathbb{F}_p^* \to \mathbb{F}_p^*,\ x \mapsto x^3$$

is a group isomorphism. With the fact that it holds $0^3 = 0$, we get that every element of $\mathbb{F}_p$ has a unique cubic root. Hence, the count of solutions to the equation

$$x^3 + y^3 + z^3 = 0$$

is equivalent to the count of solutions of the linear equation

$$x + y + z = 0.$$

This is exactly the equation of a line in the projective plane, therefore by Proposition 3.3 it has $p + 1$ rational points over $\mathbb{F}_p$, i.e. we obtain $M_p = p + 1$.

(b) The idea of this proof is first to represent M_p with $[RRR]$, then simplify this number and calculate the simplified version with the help of the p'th roots of unity. For this, let $p \equiv 1$ (mod 3), i.e. we can find $m \in \mathbb{N}$ with $p = 3m + 1$. Since 3 divides $p - 1 = |\mathbb{F}_p^*|$, the group homomorphism

$$\Phi : \mathbb{F}_p^* \to \mathbb{F}_p^*,\ x \mapsto x^3$$

is not surjective. We denote by $R = \{x^3 \mid x \in \mathbb{F}_p^*\}$ the image of Φ. We can observe that R is a subgroup of $\mathbb{F}_p^*$ of index 3, which implies that for $u \in \mathbb{F}_p^*$ with $u^3 = 1$ we obtain $\mathrm{Ker}(\Phi) = \{1, u, u^2\}$. We can conclude by saying that for any $x^3 \in R$ there are exactly $x, ux, u^2x \in \mathbb{F}_p^*$ with $x^3 = x^3, (ux)^3 = x^3$ and $(u^2x)^3 = x^3$.

We are now ready to determine the number of solutions M_p. We start by considering the simplest case, i.e. the case where one coordinate is zero. We first consider the solutions of $x^3 + y^3 + z^3 = 0$ where $z = 0$. Therefore neither x nor y can be zero because we do not count the trivial solution $(0, 0, 0)$. Hence we can choose any non-zero $x \in \mathbb{F}_p^*$ and then there are 3 possible values for y, namely $-x$, $-ux$ and $-u^2x$ which are the solutions of $y^3 = -x^3$. Therefore there are $3|\mathbb{F}_p^*| = 3(p - 1)$ solutions of $x^3 + y^3 + z^3 = 0$ with $z = 0$. By employing the same approach for $x = 0$ and then separately for $y = 0$, we can observe that there are a total of $9(p-1)$ solutions to our equation where one of the coordinates is set to zero. We divide by the $p-1$ possible multipliers since we don't distinguish proportional solutions, indeed we identify a solution (x, y, z) with all of its multiples (ax, ay, az) with $a \in \mathbb{F}_p^*$ and we can choose $|\mathbb{F}_p^*| = p-1$ available options for a. We conclude that there are

$$\frac{9(p-1)}{p-1} = 9$$

projective solutions with one coordinate zero.

Now we calculate solutions of $x^3 + y^3 + z^3 = 0$ with x, y and z non-zero. Using Definition 3.1, there are $[RRR]$ ways of writing zero as a sum of three non-zero cubes. By the discussion above, for each non-zero cube, there are 3 elements of $\mathbb{F}_p^*$ which give that cube. Thus there are $3^3[RRR] = 27[RRR]$ solutions of $x^3 + y^3 + z^3 = 0$ such that $x, y, z \neq 0$. As above, the number of projective solutions of $x^3 + y^3 + z^3 = 0$ with $x, y, z \neq 0$ is

$$\frac{27[RRR]}{p-1} = \frac{9[RRR]}{m}.$$

Combining these two results, we get that the total number of projective solutions of the equation $x^3 + y^3 + z^3 = 0$ is

$$M_p = \frac{9[RRR]}{m} + 9 = 9\left(\frac{[RRR]}{m} + 1\right).$$

We want now to count $[RRR]$ and for this we consider the two other cosets of R in $\mathbb{F}_p^*$, i.e. take $s \in \mathbb{F}_p^* \setminus R$ and let $S = sR = \{sr \mid r \in R\}$ and $T = s^2R = \{s^2r \mid r \in R\}$. Our aim is now to represent the quantity of solutions, denoted as M_p, using R, S, and T. Observe that

$$\mathbb{F}_p = \{0\} \cup R \cup S \cup T$$

is a disjoint union with $|R| = |S| = |T| = \frac{p-1}{3} = m$, hence, by Proposition 3.2 (i) we get

$$[RR\{0\}] + [RRR] + [RRS] + [RRT] = [RR\mathbb{F}_p] = |R|^2 = m^2.$$

Since $[RRS] = [sR, sR, sS] = [SST]$ and $[RRT] = [s^2R, s^2R, s^2T] = [TTS]$, we also get

$$[RR\{0\}] + [RRR] + [SST] + [TTS] = m^2. \tag{1}$$

Since $[\mathbb{F}_pTS] = |T| \cdot |S| = m^2$, we similarly have

$$[\{0\}TS] + [RTS] + [STS] + [TTS] = [\mathbb{F}_pTS] = m^2. \tag{2}$$

Note that $-1 = (-1)^3$ is a cube and therefore $R = -R$, $S = -S$ and $T = -T$. Since $(-S) \cap T = S \cap T = \emptyset$, we get that $[\{0\}TS] = 0$. Since $-R = R$, we have $[RR\{0\}] = |R| = m$. Therefore by subtracting (2) from (1), we get

$$m + [RRR] = [RTS],$$

from which it follows that

$$M_p = 9\frac{[RTS]}{m}.$$

As stated at the beginning of the proof, we are now ready to calculate $[RTS]$ with the help of the p'th roots of unity. For this let $\zeta = e^{2\pi i/p} \in \mathbb{C}$ and define

$$\alpha_1 = \sum_{r \in R} \zeta^r, \ \alpha_2 = \sum_{s \in S} \zeta^s, \ \alpha_3 = \sum_{t \in T} \zeta^t.$$

The complex numbers α_1, α_2 and α_3 are thus each a sum of m different p'th roots of unity. These complex numbers serve as tools to encode information regarding the sets R, S, and T. Specifically, they facilitate the connection between sums of elements from R, S, and T and the products of the corresponding Gauss sums. It turns out that these values are the solutions to a polynomial equation with coefficients that are integers. Our subsequent objective is to determine the specific form of this polynomial.

Claim 3.7. *$\alpha_1 + \alpha_2 + \alpha_3 = -1$.*

Proof. Since $\zeta^{p-1} + \zeta^{p-2} + ... + \zeta + 1 = \frac{\zeta^p - 1}{\zeta - 1} = 0$, we have that

$$\alpha_1 + \alpha_2 + \alpha_3 \ = \sum_{x \in R \cup S \cup T} \zeta^x = \sum_{x=1}^{p-1} \zeta^x = -1.$$

□

Claim 3.8. *$\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3 = -m$.*

Proof. We can write $\alpha_2\alpha_3$ as

$$\alpha_2\alpha_3 = \sum_{s \in S} \zeta^s \cdot \sum_{t \in T} \zeta^t \ = \sum_{s \in S, \, t \in T} \zeta^{s+t} = \sum_{x \in \mathbb{F}_p} [ST\{-x\}]\zeta^x,$$

where $[ST\{-x\}]$ is the number of pairs $(s,t) \in S \times T$ with $s + t = x$. Note that for any $r \in R$ we have

$$[ST\{-x\}] = [rS, rT, \{-rx\}] = [ST\{-rx\}]$$

which implies

$$m[ST\{-x\}] = \sum_{r \in R} [ST\{-rx\}] = [S, T, Rx] = \begin{cases} [STR] & \text{if } x \in R \\ [STS] & \text{if } x \in S \\ [STT] & \text{if } x \in T \end{cases}.$$

Define the integers a, b, c by

$$a = \frac{[STR]}{m}, \ b = \frac{[STS]}{m}, \ c = \frac{[STT]}{m}.$$

Then

$$\alpha_2\alpha_3 = \sum_{x \in \mathbb{F}_p} [ST\{-x\}]\zeta^x = \sum_{r \in R} a\zeta^r + \sum_{s \in S} b\zeta^s + \sum_{t \in T} c\zeta^t = a\alpha_1 + b\alpha_2 + c\alpha_3.$$

Similarly, we determine that

$$\alpha_1\alpha_3 = a\alpha_2 + b\alpha_3 + c\alpha_1,$$

$$\alpha_1\alpha_2 = a\alpha_3 + b\alpha_1 + c\alpha_2.$$

We conclude by noticing that

$$\begin{aligned}\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3 &= (a+b+c)(\alpha_1+\alpha_2+\alpha_3) = \\ &= -(a+b+c) = \\ &= -\left(\frac{[STR]}{m} + \frac{[STS]}{m} + \frac{[STT]}{m}\right) = \\ &= -\frac{[ST(R\cup S\cup T)]}{m} = \\ &= -\frac{[ST\mathbb{F}_p] - [ST\{0\}]}{m} = \\ &= -\frac{m^2-0}{m} = -m.\end{aligned}$$

□

Claim 3.9. *$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = 1 + 2m$.*

Proof. By Claim 3.5 and Claim 3.6 it follows that

$$\alpha_1^2 + \alpha_2^2 + \alpha_3^2 = (\alpha_1+\alpha_2+\alpha_3)^2 - 2(\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3) = 1 + 2m.$$

□

Claim 3.10. *$\alpha_1\alpha_2\alpha_3 = \frac{a+km}{3}$ with $k = 3a - m$.*

Proof. Consider the equations

$$\begin{aligned}\alpha_1(\alpha_2\alpha_3) &= \alpha_1(a\alpha_1 + b\alpha_2 + c\alpha_3), \\ \alpha_2(\alpha_1\alpha_3) &= \alpha_2(a\alpha_2 + b\alpha_3 + c\alpha_1), \\ \alpha_3(\alpha_1\alpha_2) &= \alpha_3(a\alpha_3 + b\alpha_1 + c\alpha_2).\end{aligned}$$

By adding these equations together and applying Claim 3.6 and Claim 3.7, we obtain

$$\begin{aligned}3\alpha_1\alpha_2\alpha_3 &= a(\alpha_1^2+\alpha_2^2+\alpha_3^2) + (b+c)(\alpha_1\alpha_2 + \alpha_1\alpha_3 + \alpha_2\alpha_3) = \\ &= a(1+2m) + (b+c)(-m) = \\ &= a + km,\end{aligned}$$

where $k = 2a - (b+c) = 3a - (a+b+c) = 3a - m$. □

Therefore, using these these claims we obtain that

$$M_p = 9\frac{[RTS]}{m} = 9a = 3(k+m) = 3k + p - 1,$$

and defining $A = 3k - 2$ and $B = b - c$, we finally get that

$$M_p = 3k + p - 1 = p + 1 + A.$$

The definition of A makes sense because we also have that $A \equiv 1 \pmod 3$.

We still need to prove that A and B are unique and satisfy $4p = A^2+27B^2$. We consider the polynomial $f(t) = (t-\alpha_1)(t-\alpha_2)(t-\alpha_3) \in \mathbb{Z}[t]$. Utilizing the aforementioned claims, we can simplify f and compute the square root of its discriminant by

$$\begin{aligned} f(t) &= (t-\alpha_1)(t-\alpha_2)(t-\alpha_3) \\ &= t^3 - (\alpha_1+\alpha_2+\alpha_3)t^2 + (\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3)t - \alpha_1\alpha_2\alpha_3 \\ &= t^3 + t^2 - mt - \frac{a+km}{3}. \end{aligned}$$

And

$$\begin{aligned} \sqrt{Disc_f} &= (\alpha_1-\alpha_2)(\alpha_1-\alpha_3)(\alpha_2-\alpha_3) \\ &= \alpha_2\alpha_3(\alpha_2-\alpha_3) + \alpha_1\alpha_3(\alpha_3-\alpha_1) + \alpha_1\alpha_2(\alpha_1-\alpha_2) \\ &= (a\alpha_1+b\alpha_2+c\alpha_3)(\alpha_2-\alpha_3) + (a\alpha_2+b\alpha_3+c\alpha_1)(\alpha_3-\alpha_1) \\ &\qquad + (a\alpha_3+b\alpha_1+c\alpha_2)(\alpha_1-\alpha_2) \\ &= (b-c)(\alpha_1^2+\alpha_2^2+\alpha_3^2-\alpha_1\alpha_2-\alpha_1\alpha_3-\alpha_2\alpha_3) \\ &= (b-c)(1+3m) = Bp. \end{aligned}$$

(Where $Disc_f$ denotes the discriminant of f)
We set

$$\beta_1 = 1+3\alpha_1,\ \beta_2 = 1+3\alpha_2 \text{ and } \beta_3 = 1+3\alpha_3$$

Notice that, as before with the help of our claims, we obtain

$$\beta_1+\beta_2+\beta_3 = 3(\alpha_1+\alpha_2+\alpha_3+1) = 0,$$

$$\beta_1\beta_2+\beta_1\beta_3+\beta_2\beta_3 = 3+6(\alpha_1+\alpha_2+\alpha_3)+9(\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3) = -3p,$$

$$\beta_1\beta_2\beta_3 = 1+3(\alpha_1+\alpha_2+\alpha_3)+9(\alpha_1\alpha_2+\alpha_1\alpha_3+\alpha_2\alpha_3)+27\alpha_1\alpha_2\alpha_3 = (3k-2)p = Ap.$$

In order to prove the equation that relates A and B we need to define another polynomial whose roots are $\beta_1, \beta_2, \beta_3$ with $g(t) = (t-\beta_1)(t-\beta_2)(t-\beta_3) = t^3 - 3pt - Ap$, from the three formulas just calculated. Furthermore, from the formula for the discriminant of a cubic polynomial, it follows that

$$Disc_g = -4(-3p)^3 - 27(Ap)^2 = 4\cdot 27p^3 - 27A^2p^2$$

We find that $Disc_g = 27^2 Disc_f$, since $\beta_i - \beta_j = 3(\alpha_i - \alpha_j)$ and thus we follow

$$Disc_g = 4\cdot 27p^3 - 27A^2p^2 = 27^2B^2p^2$$

and if we cancel out $27p^2$ we find

$$4p = A^2 + 27B^2.$$

Our final step is to demonstrate that A can be uniquely ascertained based on the two given conditions $4p = A^2+27B^2$ and $A \equiv 1 \pmod 3$. For this let A' and B' be integers such that $4p = A'^2 + 27B'^2$. We have

$$\begin{aligned} 4p(B'^2 - B^2) &= (A^2+27B^2)B'^2 - (A'^2+27B'^2)B^2 = \\ &= (AB'+A'B)(AB'-A'B). \end{aligned}$$

So p has to divide one of the two terms on the right-hand side of the equality. W.l.o.g. we can assume that $p \bigm| (AB' - A'B)$ (otherwise we change the sign of A'). Next we multiply the two expression for $4p$ to get

$$\begin{aligned} 16p^2 &= (A^2 + 27B^2)(A'^2 + 27B'^2) \\ &= A^2A'^2 + 27A'^2B^2 + 27A^2B'^2 + 27^2B^2B'^2 \\ &= (AA' + 27BB')^2 + 27(AB' - A'B)^2 \end{aligned}$$

With the assumption that p divides $(AB' - A'B)$, we find

$$16 - \left(\frac{AA' + 27BB'}{p}\right)^2 = 27\left(\frac{AB' - A'B}{p}\right)^2.$$

Note that the left-hand side cannot exceed 16, and the right-hand side has to be 27 times the square of an integer. Consequently, both sides have to be equal zero, implying that $AB' = A'B$ and in particular $AB' - A'B = 0$. Let $\lambda = \frac{A'}{A} = \frac{B'}{B}$, then we have $A' = \lambda A$ and $B' = \lambda B$. Replacing in

$$A'^2 + 27B'^2 = \lambda^2(A^2 + 27B^2) = \lambda^2(A'^2 + 27B'^2),$$

we obtain that $\lambda = \pm 1$. The assumption that $A \equiv A' \equiv 1 \ (mod\ 3)$ implies that $sign(A) = sign(A')$ and this forces that $\lambda = 1$.
Thus we find that A and B are distinct and we conclude the proof of Gauss' theorem. □

4 Points of Finite Order Revisited

We consider a cubic curve in Weierstrass form:

$$C\colon y^2 = x^3 + ax^2 + bx + c,$$

where $a, b, c \in \mathbb{Z}$.
Consider a finite Field $\mathbb{F}_p$ and the reduction modulo p:

$$\begin{aligned} \alpha\colon \mathbb{Z} &\to \mathbb{F}_p \\ z &\mapsto \tilde{z} \end{aligned}$$

Reducing our Curve C we obtain:

$$\tilde{C}\colon y^2 = x^3 + \tilde{a}x^2 + \tilde{b}x + \tilde{c}$$

Where $\tilde{k} = k$ mod p for $k \in \mathbb{Z}$. So we have obtained a new curve with coefficients in $\mathbb{F}_p$. Now a natural question to ask is when this curve is non-singular.

Lemma 4.1. *Let D be the discriminant of C. Then the reduced curve $\tilde{C}$ is non-singular if and only if $p \nmid D$ and $p \geq 3$*

Proof. Recall that the discriminant of a cubic polynomial of the form $x^3 + ax^2 + bx + c$ is given by:

$$D = a^2b^2 - 4a^3c - 4b^3 - 27c^2 + 18abc$$

And the discriminant of $\tilde{C}$ is given by:

$$\tilde{D} = \tilde{a}^2\tilde{b}^2 - 4\tilde{a}^3\tilde{c} - 4\tilde{b}^3 - 27\tilde{c}^2 + 18\tilde{a}\tilde{b}\tilde{c}$$

(Note that the discriminant of $\tilde{C}$ can be either calculated from directly from the expression of $\tilde{C}$ or by taking the reduction of D mod p, since the reduction modulo p from $\mathbb{Z}$ to $\mathbb{F}_p$ is a homomorphism.)
Let us first consider the case $p = 2$.
Note that the partial derivative with respect to y vanishes, that is $\frac{d}{dy}y^2 = 2y = 0$. So we can always find a singular point on $\tilde{C}$

Now consider $p \geq 3$, then we have $2y \overset{!}{=} 0$ which indicates that $y = 0$. Thus x has to be a common root of both $y^2 = x^3 + \tilde{a}x^2 + \tilde{b}x + \tilde{c}$ and its derivative and we find that $\tilde{C}$ is non-singular if and only if for the discriminant $\tilde{D}$ of $\tilde{C}$ it holds that $\tilde{D} = D \bmod p \neq 0$. That is, if p does not divide D.

□

Consider now points in $C(\mathbb{Z})$ (i.e. points in $C(\mathbb{Q})$ that happen to have integer coordinates), then we can reduce these points modulo p. Explicitly, let $P = (x, y) \in C(\mathbb{Z})$, that is, x and y satisfy

$$y^2 = x^3 + ax^2 + bx + c$$

where $a, b, c \in \mathbb{Z}$. Then we can reduce the equation modulo p and we get

$$\tilde{y}^2 = \tilde{x}^3 + \tilde{a}\tilde{x}^2 + \tilde{b}\tilde{x} + \tilde{c}$$

Which tells us that $\tilde{P} = (\tilde{x}, \tilde{y})$ is a point on $\tilde{C}(\mathbb{F}_p)$. So we can find a map from $C(\mathbb{Z})$ to $\tilde{C}(\mathbb{F}_p)$.
Recall the Nagell-Lutz theorem: We have that all points of finite order (aside from $\mathcal{O}$) in $C(\mathbb{Q})$ have integer coordinates. Consider now the following group:

$$\Phi := \{P = (x, y) \in C(\mathbb{Q}) \mid P \text{ has finite order}\} \cup \{\mathcal{O}\}$$

This is the torsion subgroup of $C(\mathbb{Q})$.
Note: Indeed this is a subgroup of $C(\mathbb{Q})$: Let P_1 and P_2 be points of finite order, then $P_1 + P_2$ and $P_1 - P_2$ are as well: Suppose $m_1P_1 = \mathcal{O}$ and $m_2P_2 = \mathcal{O}$ for some positive integers m_1 and m_2 , then we have $m_1m_2(P_1 \pm P_2) = \mathcal{O}$ as well.
Let us now define the *reduction modulo p map* on the torsion group as follows:

$$\begin{aligned} \varphi\colon \Phi &\to \tilde{C}(\mathbb{F}_p) \\ P \mapsto \bar{P} &= \begin{cases} (\bar{x}, \bar{y}) & \text{if } P = (x, y), \\ \bar{\mathcal{O}} & \text{if } P = \mathcal{O}. \end{cases} \end{aligned}$$

(We can do this, since Φ consists of points with integer coordinates.)
Note that $\tilde{C}(\mathbb{F}_p)$ is a group if $p \geq 3$ and if $p \nmid D$. In that case we have a map from a group to a group.

Theorem 4.2 (Reduction Modulo p Theorem)**.** *Let C be a non-singular cubic curve*

$$C\colon y^2 = x^3 + ax^2 + bx + c,$$

where a, b, c are integer coefficients and let D be the discriminant:

$$D = a^2b^2 - 4a^3c - 4b^3 - 27c^2 + 18abc$$

Let $\Phi \subseteq C(\mathbb{Q})$ be the subgroup consisting of all the points of finite order. Let p be a prime and let $P \to \tilde{P}$ be the reduction modulo p map

$$\begin{aligned} \varphi \colon \Phi &\to \tilde{C}(\mathbb{F}_p) \\ P \mapsto \tilde{P} &= \begin{cases} (\bar{x}, \bar{y}) & \textit{if } P = (x, y), \\ \bar{\mathcal{O}} & \textit{if } P = \mathcal{O}. \end{cases} \end{aligned}$$

If $p \nmid 2D$, then $\varphi \colon \Phi \to \tilde{C}(\mathbb{F}_p)$ is an isomorphism of Φ onto a subgroup of $\tilde{C}(\mathbb{F}_p)$.

Note that the conditions $p \geq 3$ and $p \nmid D$ are equivalent to $p \nmid 2D$.

Proof. We first show that the map φ is a homomorphism, that is, we want to show that $\widetilde{(P_1 + P_2)} = \tilde{P}_1 + \tilde{P}_2$.
We start by noticing

$$\widetilde{(-P)} = \widetilde{(x, -y)} = (\tilde{x}, -\tilde{y}) = -(\tilde{P})$$

It suffices to show that $P_1 + P_2 + P_3 = \mathcal{O}$ implies $\tilde{P}_1 + \tilde{P}_2 + \tilde{P}_3 = \tilde{\mathcal{O}}$. This suffices since if we set $P_3 = -P_1 - P_2$, then $P_1 + P_2 + P_3 = \mathcal{O}$ holds if and only if

$$P_1 + P_2 = \tilde{\mathcal{O}} - \tilde{P}_3 = \widetilde{(-P_3)} = -\tilde{P}_3 = \widetilde{(P_1 + P_2)}$$

Furthermore, if (without loss of generality) we were to set $P_3 = \mathcal{O}$, then $P_1 + P_2 + P_3 = \mathcal{O}$ would imply $P_1 = -P_2$ and thus

$$\tilde{P}_1 + \tilde{P}_2 + \tilde{P}_3 = \tilde{P}_1 + \widetilde{(-P_1)} = \tilde{P}_1 - \tilde{P}_1 = \tilde{\mathcal{O}}$$

So it suffices to show the statement for three non-trivial points which are not $\mathcal{O}$. Let these be

$$P_1 = (x_1, y_1), \quad P_2 = (x_2, y_2), \quad P_3 = (x_3, y_3)$$

Now assume that $P_1 + P_2 + P_3 = \mathcal{O}$. It follow from the definition of the group law on C that these three points have to be on a line. Let this line be $y = \lambda x + \mu$. In the case that all three points coincide, we take the tangent line.
With the explicit formula for adding points we find

$$x_3 = \lambda^2 - a - x_1 - x_2 \quad \text{and} \quad y_3 = \lambda x_3 + \mu.$$

Note that since $x_1, x_2, y_1, y_2, a \in \mathbb{Z}$, we also have $\lambda, \mu \in \mathbb{Z}$. So we can use this to reduce λ and μ modulo p. Now if we plug in the equation of the line into the equation of the cubic, we find

$$x^3 + ax^2 + bx + c - (\lambda x + \mu)^2 = 0$$

With x_1, x_2, x_3 as the three roots. So we can factorise the above equation into

$$0 = x^3 + ax^2 + bx + c - (\lambda x + \mu)^2 = (x - x_1)(x - x_2)(x - x_3)$$

Now we can reduce modulo p and find

$$x^3 + \tilde{a}x^2 + \tilde{b}x + \tilde{c} - (\tilde{\lambda}x + \tilde{\mu})^2 = (x - \tilde{x_1})(x - \tilde{x_2})(x - \tilde{x_3}).$$

And furthermore we find $\tilde{y_i} = \tilde{\lambda}\tilde{x_i} + \bar{\mu}$, for $i = 1, 2, 3$. So the line $y = \tilde{\lambda}x + \tilde{\mu}$ intersects the curve $\tilde{C}$ in the points $\tilde{P_1}$, $\tilde{P_2}$ and $\tilde{P_3}$. In the case that two points coincide, let's say (without loss of generality) $\tilde{P_1} = \tilde{P_2}$, then the line is tangent to $\tilde{C}$ at the point $\tilde{P_1}$ and similarly if all three points coincide.
So we have shown that $\tilde{P_1} + \tilde{P_2} + \tilde{P_3} = \tilde{\mathcal{O}}$ and thus φ is a group homomorphism. To see that is also an isomorphism, we note that every non-zero point (x, y) in Φ gets sent to $(\tilde{x}, \tilde{y}) \neq \tilde{\mathcal{O}}$, which means that the kernel is trivial and thus the homomorphism is one-to-one. □

Example 4.3. Consider

$$C\colon y^2 = x^3 + 3.$$

We want to find the points of finite order using the theorem above. We find $D = -3^5$, so we have $p \nmid 2D$ if and only if $p \geq 5$. So for all $p \geq 5$ there exists a one-on-one homomorphism from Φ to $\tilde{C}(\mathbb{F}_p)$. We find that $|\tilde{C}(\mathbb{F}_5)| = 6$. Specifically

$$\tilde{C}(\mathbb{F}_5) = \{\tilde{\mathcal{O}}, (\tilde{1}, \tilde{2}), (\tilde{1}, \tilde{3}), (\tilde{2}, \tilde{1}), (\tilde{2}, \tilde{4}), (\tilde{3}, \tilde{0})\}.$$

Similarly, we can also find $|\tilde{C}(\mathbb{F}_7)| = 13$. So by applying the theorem for $p = 5$ and for $p = 7$, we find that $|\Phi|$ has to divide both 6 and 13. Therefore it has to be $|\Phi| = 1$ and the only point of finite order in $C(\mathbb{Q})$ is $\mathcal{O}$.

Example 4.4.

$$C\colon y^2 = x^3 + x$$

The discriminant is $D = -2^2 = -4$. So we have a one-to-one homomorphism for $p \geq 3$. We find $|\tilde{C}(\mathbb{F}_3)| = 4$ and $|\tilde{C}(\mathbb{F}_5)| = 4$. Specifically we find:

$$\tilde{C}(\mathbb{F}_3) = \{\tilde{\mathcal{O}}, (\tilde{0}, \tilde{0}), (\tilde{2}, \tilde{1}), (\tilde{2}, \tilde{2})\}$$

$$\tilde{C}(\mathbb{F}_5) = \{\tilde{\mathcal{O}}, (\tilde{0}, \tilde{0}), (\tilde{2}, \tilde{0}), (\tilde{3}, \tilde{0})\}.$$

Note that one can actually find that $|\tilde{C}(\mathbb{F}_p)|$ is divisible by 4 for all primes $p \geq 3$. Using the fact that a point in $\tilde{C}$ has order 2 if and only if its y coordinate is zero, we find that

$$\tilde{C}(\mathbb{F}_3) \cong \mathbb{Z}/4\mathbb{Z} \quad \text{and} \quad \tilde{C}(\mathbb{F}_5) \cong \mathbb{F}_2 \oplus \mathbb{F}_2.$$

Since Φ has to be isomorphic to a subgroup of both of these two groups, Φ has to either be cyclic of order two or be trivial. We note that $(0, 0)$ in $C(\mathbb{Q})$ has order two (since it's y-coordinate is 0) and we conclude $\Phi = \{\mathcal{O}, (0, 0)\}$.

Example 4.5.

$$C\colon y^2 = x^3 - 43x + 166$$

The discriminant is $D = -425984 = -2^{15} * 13$ We start by applying the Nagell-Lutz theorem and find the point $P = (3, 8)$. By using the duplication formula we find the x-coordinates:

$$x(P) = 3 \quad x(2P) = -5 \quad x(4P) = 11 \quad \text{and} \quad x(8P) = 3$$

Note that $x(8P) = x(P)$, so we have $8P = \pm P$. Furthermore, we have that $2D$ is relatively prime to 3, so we know that Φ has to be a subgroup of $\tilde{C}(F_3)$, for which we find $|\tilde{C}(F_3)| = 7$. So Φ has to have order 1 or 7. Since we already know that Φ contains the point P, we conclude that the points of finite order in $C(\mathbb{Q})$ form a cyclic group of order 7 generated by $P = (3, 8)$. We find specifically

$$\Phi = \{\mathcal{O}, (3, \pm 8), (-5, \pm 16), (11, \pm 32)\}$$

5 Reduction of an elliptic curve

Let us consider an elliptic curve

$$E\colon Y^2Z = X^3 + aXZ^2 + bZ^3,$$

with $a, b \in \mathbb{Q}$ and $D = 4a^3 + 27b^2 \neq 0$. By changing the variables, we can minimize $|D|$. To do this we set $X \mapsto X/c^2$ and $Y \mapsto Y/c^3$ such that the new a and b are integers and $|D|$ is minimal. The new equation is then said to be *minimal.*

Now let us consider

$$\tilde{E}\colon Y^2Z = X^3 + \tilde{a}XZ^2 + \tilde{b}Z^3$$

where $\tilde{a}$ and $\tilde{b}$ are the images of a and b in $\mathbb{F}_p$. We call $\tilde{E}$ the *reduction of E modulo p.* There are different cases for us to consider:

- *Good reduction* If $p \neq 2$ and p does not divide D (equivalent to $p \nmid 2D$), then $\tilde{E}$ is an elliptic curve over $\mathbb{F}_p$ and we call this reduction *good.*
- *Cuspidal or additive reduction* If $p \neq 2, 3$, $p \mid D$ and $p \mid (-2ab)$. Then $\tilde{E}$ has a cusp and we say the reduction is *cuspidal.*
- *Nodal or multiplicative reduction* If $p \neq 2, 3$, $p \mid D$ and $p \nmid (-2ab)$. Then $\tilde{E}$ has a node and the reduction is called *nodal.* In this case we have two sub-cases:
  - *Split multiplicative reduction* If $-2ab$ is a square modulo p. In this case the tangents at the node are rational over $\mathbb{F}_p$.
  - *Nonsplit multiplicative reduction* If $-2ab$ is not a square modulo p. In this case the tangents at the node are not rational over $\mathbb{F}_p$

Note that for cases $p = 2$ and $p = 3$ we need to consider the full equation

$$Y^2Z + a_1XYZ + a_3YZ^2 = X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3$$

since it might be possible to find a "more minimal" equation of this form. As before, a n equation for E would be considered *minimal* if all a_i are integers and $|D|$ is minimal. An example would be

$$Y^2 + Y = X^3 - X^2$$

which defines a non-singular curve over $\mathbb{F}_p$. But every equation of the form $Y^2Z = X^3 + aXZ^2 + bZ^3$ would define a singular curve over $\mathbb{F}_2$.

References

[1] J.H. Silverman and J.T. Tate, *Rational Points on Elliptic Curves*, Springer (1992).

[2] J.S. Milne, *Elliptic Curves*, BookSurge Publishers (2006).