1. The cubic curve.

Let k be a field, let $a_1, a_2, a_3, a_{4,6} \in k$. Let E denote the projective curve given in $\mathbf{P}_2$ by the homogeneous cubic equation

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

The line given by the equation $Z = 0$ is the line at infinity. It intersects E in a unique point namely $(0 : 1 : 0)$. We denote this point by ∞. The goal of this note is to give a short self-contained proof of the well-known fact that the chord and tangent proces equips the set $E(k)$ of non-singular k-rational points of E with a group structure whose neutral element is the point ∞.

On the affine chart given by $Y = 1$, the equation of E becomes

$$Z + a_1XZ + a_3Z^2 \;=\; X^3 + a_2X^2Z + a_4XZ^2 + a_6Z^3.$$

This shows that ∞ is a non-singular point of E. It is a flex point and its tangent line is precisely the line at infinity. All other points of E have their Z-coordinates different from zero. They are precisely the points that of E that are on the affine chart given by $Z = 1$. On this chart the equation of E becomes the usual inhomogenous Weierstrass equation

$$Y^2 + a_1XY + a_3Y \;=\; X^3 + a_2X^2 + a_4X + a_6.$$

We denote this affine curve by E^0. The set of k-rational points $E^0(k)$ of E^0 is equal to $E(k)$ minus the point $(0 : 1 : 0)$. The lines given by $X = \alpha Z$ in $\mathbf{P}_2$ all pass through $(0 : 1 : 0)$. Their intersections with the affine chart given by $Z = 1$ are vertical lines given by equations of the form $X = \alpha$.

Putting $f(X, Y) = -Y^2 - a_1XY - a_3Y + X^3 + a_2X^2 + a_4X + a_6$, the relevant ring is

$$R = k[X, Y]/(f(X, Y)).$$

It is a quadratic extension of the polynomial ring $k[X]$. It is a domain since the degree 2 polynomial $f(X, Y) \in k(X)[Y]$ has no zeroes in $k(X)$.

Let $P = (\alpha, \beta)$ be a point in $E(k)$ different from ∞. The map $g(X, Y) \mapsto g(\alpha, \beta)$ is a k-algebra morphism $R \longrightarrow k$. Its kernel is the maximal ideal $\mathfrak{m}_P = (X - \alpha, Y - \beta)$. Conversely, every maximal ideal $\mathfrak{m}$ for which $R/\mathfrak{m} \cong k$ is of this form.

Proposition 1.1. *Let P be a nonsingular point in $E^0(k)$ and let $\mathfrak{m}$ be the correspdonding maximal ideal of R. Let $\ell \in R$ be such that $\ell = 0$ is an equation for the tangent line at P. Then ℓ is in $\mathfrak{m}^2$ and $\mathfrak{m}/\mathfrak{m}^2$ is a 1-dimensional k-vector space. We have $\ell \in \mathfrak{m}^3$ if and only if P is a flex point of E*

Proof. Let $P = (\alpha, \beta)$. We the variables X and Y by $X + \alpha$ and $Y + \beta$. In terms of the new coordinates the Weierstrass equation becomes

$$Y^2 + a_1XY + a_3Y \;=\; X^3 + a_2X^2 + a_4X.$$

So we have $a_6 = 0$ and $R/\mathfrak{m}^2 = k[X,Y]/(X^2, XY, Y^2, a_3Y - a_4X)$. This shows that $\ell = a_3Y - a_4X$ is in $\mathfrak{m}^2$. Since P non-singular ℓ is not zero and the k-dimension of $\mathfrak{m}/\mathfrak{m}^2$ is 1.

If a_3 is zero, the intersection of E^0 with ℓ is given by the equation $Y^2 = 0$, so that P cannot be a flex. So, assume $a_3 = 0$ and replace ℓ by $Y - \lambda X$ with $\lambda = a_4/a_3$. Substituting $Y = \lambda X$ in the cubic equation we obtain

$$\ell(a_3 + \varepsilon) \;=\; X^2(\lambda^2 + a_1\lambda - a_2) - X^3,$$

for some ε in the ideal (X,Y). Since $a_3 \neq 0$ it follows that ℓ is in the cube of the ideal $(X,Y) = \mathfrak{m}_P$ if and only the coefficient $\lambda^2 + a_1\lambda - a_2$ vanishes. This happens precisely when P is a flex point, as required.

2. Degrees.

We adopt the notation of section 1. The ring R is free of rank 2 over $k[X]$. More precisely, for every element $g(X,Y)$ of R there are unique polynomials $a(X), b(X) \in k[X]$ for which $g(X,Y) = a(X) + b(X)Y$. We let

$$\overline{Y} \;=\; -Y - a_1X - a_3$$

and we write $\overline{g}$ for the element $g(X,\overline{Y})$ of R. The map $g \mapsto \overline{g}$ is a $k[X]$-linear involution of R. For any element $g \in R$ we define its *norm* by

$$N(g) = \overline{g}g.$$

Proposition 2.1.
(a) The norm is a multiplicative function $R \longrightarrow k[X]$.
(b) For $g = a(X) + b(X)Y$ the polynomial $N(g)$ has degree $\max(2\text{degree}(a(X)), 3 + 2\text{degree}(b(X)))$.
(c) The unit group of R^* is equal to k^*.
(d) The degree of $N(g)$ of an element $g \in R - k$ is at least 2.

Proof. Part (a) is clear. Part (b) follows from the shape of the Weierstrass equation. Any $g \in R$ is invertible if and only if $N(g)$ is invertible in $k[X]$. Part (c) follows therefore from the fact that the unit group of the ring $k[X]$ is k^*. Part (d) follows from (b) and (c).

Any non-zero ideal $I \subset R$ contains a non-zero polynomial in $k[X]$ and has therefore finite codimension in R. The *degree* $\deg I$ of I is $\dim_k R/I$.

Lemma 2.2. *Let $I \subset R$ a non-zero ideal and let g be a non-zero element of R. Then*

$$\deg gI \;=\; \deg I + \deg gR.$$

Proof. This follows from the exactness of the sequence

$$0 \longrightarrow R/I \xrightarrow{g} R/gI \longrightarrow R/gR \longrightarrow 0.$$

Clearly we have $\deg R = 0$. A direct computation shows that the ideals XR and YR have degrees 2 and 3 respectively.

Proposition 2.3. *Let $g \in R$ be a non-zero element. Then*

$$\deg gR \;=\; \text{degree}\, N(g).$$

Here "degree" denotes the usual degree of a polynomial in $k[X]$.

Proof. The involution induces a ring isomorphism $R/gR \cong R/\overline{g}R$. This shows that $\deg gR = \deg \overline{g}R$. Since R is free of rank 2 over $k[X]$, for any polynomial $p \in k[X] \subset R$ we have $\deg pR = 2\text{degree}\, p$. Now apply Lemma 2.2 to $I = \overline{g}R$. Since $\overline{g}g$ is in $k[X]$ we get

$$2\deg gR = \deg gR + \deg \overline{g}R = \deg \overline{g}gR = 2\text{degree}\, \overline{g}g$$

as required.

For $i \geq 0$ we put $e_{2i} = X^i$ and $e_{2i+1} = YX^{i-1}$. Then $e_1 = 1$ has degree 0 while $\deg e_k = k$ for all $k \geq 2$. There are no elements of degree 1 in R.

Proposition 2.4. *(Lenstra-Riemann-Roch) Let $I \subset R$ be an ideal of finite codimension. Then there is an element $a \in I$ for which the ideal aR has codimension ≤ 1 in I. Moreover, a is unique up to multiplication by k^*.*

Proof. Let $d = \dim R/I$. Then the elements $e_1, \ldots, e_{d+1}$ are dependent in the k-vector space R/I. Take for a any non-zero linear combination that is in I. The result now follows from the fact that the degree of a linear combination of $e_1, \ldots, e_{d+1}$ is at most $d+1$.

If there is an element $a \in I$ for which the codimension is 0, then a actually generates I. Therefore it has degree d and is unique up to multiplication by $R^* = k^*$. For any $b \in I$ of degree $d+1$, the element b/a has degree 1, which is impossible. Therefore there are no elements of degree $d+1$ and a is unique.

If there is no element for which the codimension is 0, then the element a for which the codimension is 1 is unique up to multiplication by elements of k^*. Indeed, if there were two elements a that are not k-multiples of one another, then they would both have degree $d+1$ and a suitable k-linear combination b would be a non-zero element of degree $\leq d$ and therefore generate I, contradiction.

3. Invertible ideals.

Let A be a domain. An ideal $I \subset A$ is called *invertible* if there exists an ideal $J \subset A$ for which the product IJ is a non-zero principal ideal. Clearly non-zero principal ideals are invertible. Two invertible ideals I and I' are called *equivalent* if there exist non-zero elements $a, a' \in A$ for which $aI = a'I'$. It is easy to see that this is an equivalence relation. All non-zero principal ideals are equivalent to one another. Ideal multiplication turns the set of equivalence classes into a group whose neutral element is the class of principal ideals. This group is the *class group* of R. It is denoted by $Cl(R)$.

We apply this to the ring R of the previous section. For every k-point $P = (\alpha, \beta)$ of the curve E given by the Weierstrass equation

$$Y^2 + a_1XY + a_3Y \;=\; X^3 + a_2X^2 + a_4X + a_6,$$

the map $g(X,Y) \mapsto g(\alpha, \beta)$ is a k-algebra morphism $R \longrightarrow k$. Its kernel is the maximal ideal $\mathfrak{m}_P = (X - \alpha, Y - \beta)$. It has residue field k. Conversely, every k-algebra morphism $R \longrightarrow k$ is of this form.

Proposition 3.1. *For every non-singular k-point P of E, the corresponding ideal $\mathfrak{m}_P$ satisfies $\mathfrak{m}_P\overline{\mathfrak{m}}_P = (X-\alpha)R$ and is therefore invertible. Moreover, if $P = \overline{P}$, the tangent line of E to P is given by the equation $X = \alpha$.*

Proof. Let $P = (\alpha, \beta)$. If $\mathfrak{m}_P$ is not stable under the involution of R, then $\mathfrak{m}_P$ and $\overline{\mathfrak{m}}_P = \mathfrak{m}_{\overline{P}}$ are coprime maximal ideals. Here $\overline{P}$ denotes the points $(\alpha, -\beta - a_1\alpha - a_3)$. By the Chinese Remainder Theorem, the ideal $\mathfrak{m}_P\overline{\mathfrak{m}}_P = \mathfrak{m}_P \cap \overline{\mathfrak{m}}_P$ has codimension 2. Since it contains the degree 2 element $X - \alpha$, we have the equality $\mathfrak{m}_P\overline{\mathfrak{m}}_P = (X-\alpha)R$ and we are done.

If $\mathfrak{m}_P = \overline{\mathfrak{m}}_P = \mathfrak{m}_{\overline{P}}$ we translate P to the origin $(0,0)$. This means that we replace $X, Y \in R$ by $X + \alpha$ and $Y + \beta$ respectively. In terms of these new coordinates, $\overline{P}$ becomes the point $(0, -a_3)$. Since $P = \overline{P}$ we must have $a_3 = 0$. The equation of E now becomes

$$Y^2 + a_1XY = X^3 + a_2X^2 + a_4X.$$

Since P is not singular, the coefficient a_4 is not zero and hence the tangent line to P is given by $X = 0$. Since the degree 2 ideal $\mathfrak{m}_P^2$ contains the function X, it is in fact generated by it. In terms of the old coordinates this means precisely $\mathfrak{m}_P\overline{\mathfrak{m}}_P = (X-\alpha)R$. Moreover, the tangent line at P is given by $X = \alpha$.

This proves the proposition.

Lemma 3.2. *Let J be an invertible R-ideal. Suppose $I \subset I' \subset R$ are ideals for which $\dim_k I/I' = 1$. Then also $\dim_k IJ/I'J = 1$.*

Proof. Let $e \in I - I'$, then the kernel of the surjective morphism $R \longrightarrow I/I'$ given by $\lambda \mapsto \lambda e$ is a maximal ideal $\mathfrak{m}$ of R with residue field k. The quotient $IJ/I'J$ is an $R/\mathfrak{m}$-vector space. Its subspaces are of the form $\mathfrak{a}/I'J$ for some R-ideal $\mathfrak{a}$ between $I'J$ and IJ. Since J is invertible, there exists an ideal $K \subset R$ for which $JK = (b)$ for some non-zero $b \in R$. Multiplication by K is an injective map

$$\begin{array}{c}\{\text{ideals between } I'J \text{ and } IJ\}\\ \downarrow \\ \{\text{ideals between } bI \text{ and } bI'\}\end{array}$$

Since bI/bI' is isomorphic to I/I', its k-dimension is 1. Therefore there are no ideals properly between bI' and bI. It follows that there are no ideals properly between $I'J$ and IJ either. This implies that the k-vector space $IJ/I'J$ has dimension 1, as required.

Proposition 3.3. *Let $I \subset R$ be invertible. Then either I is principal, or there is a unique k-point P of E for which*

$$I\mathfrak{m}_P = aR.$$

for some non-zero $a \in I$ that is unique up to multiplication by k^*.

Proof. By Proposition 2.3 there is an element $a \in I$ such that the codimension of aR in I is at most 1. If a generates I, we are done. Suppose that the codimension is 1. Since I is invertible there is an ideal J for which $IJ = (b)$ for some non-zero $b \in R$. By Lemma 3.2 the codimension of aJ in bR is 1. It follows that the $(a/b)J$ an R-ideal is of codimension 1. It is therefore of the form $\mathfrak{m}_P$ for some k-point P. Moreover, since I is invertible, the point P is unique.

Corollary 3.4. *The map* $E(k) \longrightarrow Cl(R)$ *given by*

$$\begin{array}{rcll} \infty & \mapsto & R, & \\ P & \mapsto & \mathfrak{m}_P, & \text{for } P \neq \infty, \end{array}$$

is a bijection.

Proof. Surjectivity follows from Proposition 3.3. It remains to show that the map is injective. Since there are no elements of degree 1 in R, no ideal $\mathfrak{m}_P$ is principal and therefore only the point ∞ is being mapped to the neutral element of $Cl(R)$. Suppose that for two points $P, Q \in E(k) - \{\infty\}$ we have $a\mathfrak{m}_P = b\mathfrak{m}_Q$ for certain non-zero $a, b \in R$. It follows that $\mathfrak{m}_P\mathfrak{m}_{\overline{Q}}$ is principal. By Lemma 3.2 its codimension is 2. So it is generated by an element of the form $X - \alpha$ for some $\alpha \in k$. Proposition 3.1 says that $\mathfrak{m}_P\mathfrak{m}_{\overline{P}} = (X - \alpha)$. Therefore we have $\mathfrak{m}_{\overline{P}} = \mathfrak{m}_{\overline{Q}}$ and hence $P = Q$ as required.

The bijection of Corollary 3.4 transports the group structure of the class group $Cl(R)$ to the set $E(k)$. In this way we find that the set $E(k)$ of non-singular k-rational points of E becomes a group with neutral element the point ∞.

In the rest of this section we check that the group operations agree with the usual chord and tangent procedures. First of all, the point at infinity is the neutral element for the chord and tangent composition. Second, for any point $P = (\alpha, \beta)$ in $E^0(k)$ the vertical line through P has precisely two k-rational points of intersection with the affine cubic curve. This follows from the fact that substituting $X = \alpha$ in the Weierstrass equation leads to a quadratic equation in Y one of whose roots is α. The second point is $\overline{P}$. Proposition 3.1 says that $\mathfrak{m}_P\mathfrak{m}_{\overline{P}} = (X - \alpha)R$. This agrees with the fact that P and $\overline{P}$ are opposite points for the chord and tangent composition.

Finally, suppose that P and Q are points of E that are not at infinity and for which $P \neq \overline{Q}$. Then the line passing through P and Q, or the tangent line at P if $P = Q$, is not vertical and has an equation of the form $Y = \lambda X + \mu$ for certain $\lambda, \mu \in k$. The line has precisely three k-rational points of intersection with the affine cubic curve. This follows from the fact that substituting $Y = \lambda X + \mu$ in the Weierstrass equation leads to a cubic equation in X. If $P \neq Q$, the X-coordinates of P and Q are distinct zeroes and there is a unique third point of intersection R. If $P = Q$, we translate P to the point $(0, 0)$. The tangent line at P has equation $a_3Y = a_4X$ with $a_3 \neq 0$. Putting $\lambda = a_4/a_3$ and substituting Y by λX in the Weierstrass equation, leads to a cubic polynomial divisible by X^2. Therefore there is also a unique third point of intersection R in this case.

In order to show that the composition with chord and tangents agrees with the group law in $Cl(R)$ it suffices to show that

$$\mathfrak{m}_P\mathfrak{m}_Q\mathfrak{m}_R \;=\; (Y - \lambda X - \mu)R.$$

By Propositions 2.3 and 3.1 the degrees of the ideals $(Y - \lambda X - \mu)R$ and $\mathfrak{m}_P\mathfrak{m}_Q\mathfrak{m}_R$ are 3. Therefore it suffices to show that the element $Y - \lambda X - \mu$ is contained in the product of the three maximal ideals. If P, Q and R are all distinct, this is clear by the Chinese Remainder Theorem. If, say, $P = Q$ but $P \neq R$, then Proposition 1.1. implies that $Y - \lambda X - \mu$ is contained in $\mathfrak{m}_P^2$ and $\mathfrak{m}_R$ and therefore in their product. Finally, if $P = Q = R$, we have a flex point and Proposition 1.1 implies that $Y - \lambda X - \mu$ is in $\mathfrak{m}_P^3$ as required.

4. The zeta function.

Let $k = \mathbf{F}_q$ be a finite field and let E be an elliptic curve defined over k. This means that E a cubic plane curve as in section 1, that is non-singular. Let R be the ring $\mathbf{F}_q[X, Y]$ modulo the Weierstrass equation. The zeta function of R is defined by

$$Z_R(T) \;=\; \sum_{I \subset R} T^{\deg I}, \quad \in \mathbf{Z}[[T]].$$

Here I runs over the non-zero ideals of R. Note that for each d there are only finitely many ideals of degree d. We count them in the proof of the next proposition.

Proposition 4.1. *We have*

$$Z_R(T) \;=\; \frac{1 - (h-q)T + qT^2}{1 - qT},$$

where h is the number of degree 1 ideals.

Proof. The ring R itself is the only ideal of degree 0. Let $d \geq 2$. The space L_d of elements $f \in R$ of degree d has $q^d - q^{d-1}$ elements. Therefore there are q^{d-1} principal ideals of degree d. If I is a non-principal ideal of degree d, then Proposition 3.3 implies that there is an element $f \in I$ of degree $d+1$ and a codimension 1 maximal ideal $\mathfrak{m}_P$ for which

$$(f) \;=\; I\mathfrak{m}_P.$$

Moreover, the ideal $\mathfrak{m}_P$ is unique and f is unique up to units. The subspaces of the functions that vanish in P have codimensions 1 in L_{d+1} and L_d respectively. Therefore, up to units of R, there are q^{d-1} possibilities for f. Since there are h possibilities for the points P, it follows that there are $(h+1)q^{d-1}$ ideals of degree d.

Therefore we have

$$Z_R(T) \;=\; 1 + hT + \sum_{d=2}^{\infty} (h+1)q^{d-1}T^d \;=\; \frac{1 + (h-q)T + qT^2}{1 - qT}$$

as required.

The zeta function $Z_R(T)$ is not equal to the usual zeta function $Z_E(T)$ of the elliptic curve E. But it is closely related. In order to explain the connection, we recall the definition of $Z_E(T)$. A *place* of E is a Galois conjugacy classes of points. The degree of a place is the degree of the residue field of any of its points. So, places of degree 1 are simply k-rational points. *Divisors* of E are finite formal sums of the form

$$D \;=\; \sum_P n_P P,$$

where P runs over the places of E and the coefficients $n_P \in \mathbf{Z}$ are almost all zero. The degree of a D is equal to $\sum_P n_P \deg P$. A divisor is called *effective* if $n_P \geq 0$ for all places P. The zeta function of E is defined by

$$Z_E(T) \;=\; \prod_P \frac{1}{1 - T^{\deg P}}.$$

Here P runs over the places of E. The zeta function $Z_{E^0}(T)$ of the affine curve E^0 is given by the same product, omitting the point $P = \infty$. Writing the factors on the right as geometric series in $\mathbf{Z}[T]]$, we see that the zeta function of E is equal to the power series

$$Z_E(T) \;=\; \sum_D T^{\deg D}, \qquad \in \mathbf{Z}[T]].$$

Here D runs over the effective divisors of E. Similarly, the zeta function of E^0 is equal to

$$Z_{E^0}(T) \;=\; \sum_D T^{\deg D}, \qquad \in \mathbf{Z}[T]],$$

where D runs over the effective divisors of E^0. In other words, D runs over effective divisors of the form $\sum_P n_P \deg P$, with $n_\infty = 0$.

Theorem 4.2. *The zeta functions $Z_R(T)$ and $Z_{E^0}(T)$ are equal.*

In order to prove the theorem, we extend the correspondence between and maximal R-ideals of codimension 1 and points $P \in E(k)$ to arbitrary maximal R-ideals. A point $P = (\alpha, \beta)$ with coordinates in $\overline{k}$ corresponds to the kernel of the k-algebra morphism $R \longrightarrow \overline{k}$ that maps X and Y to α and β respectively. Conversely, any maximal ideal $\mathfrak{m}$ of R is the kernel of a k-algebra homomorfism $R \longrightarrow R/\mathfrak{m} \hookrightarrow \overline{k}$. The point (α, β) is unique up to conjugation by $\mathrm{Gal}(\overline{k}/k)$. So, apart from the point ∞, places correspond bijectively to maximal R-ideals.

The following property of the degree is useful.

Lemma 4.2. *Let $I, J \subset R$ be invertible ideals. Then*

$$\deg I + \deg J \;=\; \deg IJ.$$

Proof. If I is principal, this follows from Lemma 3.1. If not, we have $I\mathfrak{m}_P = aR$ for some non-zero $a \in I$. By Lemma 3.2. aJ has codimension 1 in IJ. Therefore Lemma 3.1 implies

$$\deg IJ = \deg aJ - 1 = \deg aR + \deg J - 1 = \deg I + \deg J.$$

as required.

Proposition 4.3. *Every non-zero ideal $I \subset R$ is a product of invertible maximal ideals in a unique way.*

Proof. Let I be an invertible R-ideal. Then $I \subset \mathfrak{m}$ for some maximal ideal. Since all points of E are non-singular, the ideal $\mathfrak{m}$ is invertible. Let $J \subset R$ be an ideal for which $J\mathfrak{m} = aR$ for some non-zero $a \in R$. Then we have $IJ \subset aR$ so that $I' = IJ/a$ is an invertible R-ideal. We have

$$I'\mathfrak{m} \;=\; IJ\mathfrak{m}/a \;=\; I.$$

By Lemma 4.2 the degree of I' is strictly smaller than $\deg I$. Repeating this construction with I' rather than I, we eventually find that I is a product of maximal R-ideals. The uniqueness follows from the fact that all maximal ideals are invertible.

This proves the lemma.

Proof. Proposition 4.3. implies that the correspondence between places different from ∞ and maximal ideals of R extends to a natural bijection between effective divisors of the affine curve E^0 and invertible R-ideals. Since the degree of $I \subset R$ is equal to the degree of the corresponding effective divisor, the zeta functions $Z_R(T)$ and $Z_E(T)$ are equal, as required.

Let $1 - (h-q)T + qT^2$ be the numerator of the zeta function of E and let π and π' be the complex zeroes of the reciprocal polynomial $T^2 - (h-q)T + q$.

Proposition 4.4. *For every $d \geq 1$, we have*

$$\#E^0(\mathbf{F}_{q^d}) \;=\; q^d - \pi^d - \pi'^d.$$

Proof. By Proposition 5.1 we have

$$Z_{E^0}(T) = \frac{1 - (h-q)T + qT^2}{1 - qT}.$$

Combining this with the product formula of Corollary 5.3 we obtain

$$Z_{E^0}(T) \;=\; \frac{(1-\pi T)(1-\pi' T)}{1-qT} = \prod_P \frac{1}{1 - T^{\deg P}}.$$

Here the product runs over the places of E^0. Taking the logarithmic derivative of this identity, expanding the geometric series and comparing coefficients shows that for every $d \geq 1$ we have $q^d - \pi^d - \pi'^d = \sum_P \deg P$ where P runs over the places of E^0 of degree dividing d. Since the latter sum is equal to $\#E^0(\mathbf{F}_{q^e})$, the lemma follows.

5. An upper bound.

In this section we obtain an upper bound for the number of points of an elliptic curve E over a finite field. Our method is due to S.A. Stepanov.

Let $\mathbf{F}_q$ denote a field of cardinality q. Let E be an elliptic curve given by a Weierstrass equation and as before let R be the $\mathbf{F}_q$-algebra generated by the functions X and Y. For $d \geq 0$ let L_d denote the $\mathbf{F}_q$-vector space

$$L_d = \{f \in R : \deg f \leq d\}.$$

As explained in section 1, we have $L_d = 0$ for $d = 0$, while for $d \geq 1$, the monomials e_i with $i \leq d$ are an $\mathbf{F}_q$-basis for L_d. In particular, L_a has $\mathbf{F}_q$-dimension d.

For $d \geq 1$ the set $L_d^q = \{f^q : f \in L_d\}$ is an $\mathbf{F}_q$-vector space of dimension $d = \dim L_d$. Indeed, the map $f \mapsto f^q$ is an $\mathbf{F}_q$-linear bijection $L_d \leftrightarrow L_d^q$.

Lemma 5.1. *Let $a, b \geq 1$ and let $L_a^q L_b$ denote the $\mathbf{F}_q$-vector space generated by the functions $f^q g$ where $f \in L_a$ and $g \in L_b$. Then we have*
(a) $\dim L_a^q L_b \leq aq + b$;
(b) $\dim L_a^q L_b \leq ab$;
(c) if $b < q$, the elements $e_i^q e_j$ for $1 \leq i \leq a$ and $1 \leq j \leq b$ form an $\mathbf{F}_q$-basis of $L_a^q L_b$ and we have equality in (b).

Proof. Part (a) follows from the fact that $L_a^q L_b \subset L_{aq+b}$. The inequality of part (b) follows from the fact that the functions $e_i^q e_j$ with $1 \leq i \leq a$ and $1 \leq j \leq b$ generate $L_a^q L_b$. For (c) we observe that

$$\deg e_i^q e_j = q \deg e_i + \deg e_j$$

So if $b < q$, we have $\deg e_j < q$ for all j. It follows that the degrees $\deg e_i^q e_j$ are all distinct. So any $\mathbf{F}_q$-linear combination $\sum_{i,j} \lambda_{ij} e_i^q e_j$ that is zero, necessarily has $\lambda_{ij} = 0$ for every i, j. This proves that the functions $e_i^q e_j$ are independent. So they are a basis for $L_a^q L_b$. This proves the lemma.

From now on we assume that $a, b \geq 1$ with $b < q$. Lemma 5.1 (c) implies that the $\mathbf{F}_q$-linear map

$$\vartheta : L_a^q L_b \longrightarrow L_a L_b^q$$

given by

$$e_i^q e_j \mapsto e_i e_j^q, \qquad \text{for } 1 \leq i \leq a \text{ and } 1 \leq j \leq b,$$

is well defined.

The following proposition is the key ingredient in the proof of Theorem 5.3.

Proposition 5.2. *Let $a, b \geq 1$ with $b < q$. If the map ϑ is not injective, then*

$$\#E(\mathbf{F}_{q^2}) \leq aq + b + 1.$$

Proof. Every function $F \in \ker \vartheta$ vanishes on $E(\mathbf{F}_{q^2}) - \{\infty\}$. Indeed, let $F = \sum \lambda_{ij} e_i^q e_j$ for certain $\lambda_{ij} \in \mathbf{F}_q$ and let $P \in E(\mathbf{F}_{q^2}) - \{\infty\}$. Then

$$F(P)^q = \sum \lambda_{ij} e_i^{q^2}(P) e_j^q(P) = \sum \lambda_{ij} e_i(P) e_j^q(P) = (\sum \lambda_{ij} e_i e_j^q)(P) = \vartheta(F)(P) = 0,$$

which is zero when $F \in \ker \vartheta$. The second equality follows from the fact that $P \in E(\mathbf{F}_{q^2})$ so that $f^{q^2}(P) = f(P)$ for every function $f \in R$.

Since ϑ is not injective, there exists a non-zero F in $\ker \vartheta$. Therefore we obtain the following estimate.

$$\#E(\mathbf{F}_{q^2}) - 1 \leq \#\{\text{zeroes of } F\} = \deg(F) \leq aq + b.$$

The rightmost inequality follows from Lemma 5.1 (a). This proves the proposition.

Theorem 5.3. *Let E be an elliptic curve defined over $\mathbf{F}_q$ and suppose that $q \geq 5$. Then we have*

$$\#E(\mathbf{F}_{q^2}) \leq q^2 + 3q.$$

Proof. The map ϑ defined above cannot be injective if $a, b \geq 1$ have the property that

$$\dim L_a^q L_b \ > \ \dim L_a L_b^q.$$

Since $b < q$, Lemma 5.1 (b) implies that $L_a^q L_b$ has dimension ab. Lemma 3.1 (b) cannot be applied to $L_a L_b^q$. In some sense this is the point of the proof. But by Lemma 3.1 (a) we know that $L_a L_b^q$ has dimension $\leq a + bq$. Therefore the map ϑ is *not* injective when

$$ab \ > \ a + bq.$$

In order to deduce a sharp estimate from Proposition 5.2, we choose a as small as possible. Since the inequality $ab > a + bq$ must be satisfied, the minimal choice is $a = q + 2$. Once a is chosen, we can take $b = q - 1$, at least for $q \geq 5$. With these choices the quantity $aq + b + 1$ in Proposition 3.3 becomes $(q+2)q + q - 1 + 1 = q^2 + 3q$, as required.

6. Hasse's Theorem.

In this section we prove Hasse's Theorem. It is the analogue of the Riemann Hypothesis for the zeta function of E. Let E be an elliptic curve over $\mathbf{F}_q$. First we use Stepanov's method to obtain a *lower* bound for $\#E(\mathbf{F}_{q^2})$ as follows.

Proposition 6.1. *Let E be an elliptic curve over $\mathbf{F}_q$ and suppose that $q \geq 5$. Then we have*

$$\#E(\mathbf{F}_{q^2}) \ > \ q^2 - 3q$$

Proof. Let Ω denote the set of points (x, y) of $E^0(\overline{\mathbf{F}}_q)$ for which $x \in \mathbf{F}_{q^2}$. For every $x \in \mathbf{F}_{q^2}$ there are at most two points $(x, y) \in \Omega$. If (x, y) is one such point, then $(x, \overline{y})$ where $\overline{y} = -y - a_1 x - a_3$, is the other. We have

$$\#\Omega \ = \ 2q^2 - r.$$

where r is the number of values of x for which $y = \overline{y}$. We have $r \leq 3$.

The automorphism σ of $\overline{\mathbf{F}}_q$ given by $\sigma(t) = t^{q^2}$ also acts on Ω. It maps a point $(x, y) \in \Omega$ to $(\sigma(x), \sigma(y)) = (x^{q^2}, y^{q^2}) = (x, y^{q^2})$. It follows that either $\sigma(y) = y$ or $\sigma(y) = \overline{y}$. Therefore have

$$\Omega = \Omega^+ \cup \Omega^-,$$

where $\Omega^+ = \{(x, y) \in \Omega : \sigma(y) = y\}$ and $\Omega^- = \{(x, y) \in \Omega : \sigma(y) = \overline{y}\}$. The intersection $\Omega^+ \cap \Omega^-$ consists of the r points (x, y) for which $y = \overline{y}$.

Clearly Ω^+ is the set $E(\mathbf{F}_{q^2}) - \{\infty\}$. Theorem 3.4 provides an estimate for its size. In this section we use the method of section 3 to obtain an estimate of the size of the set Ω^-. Let a, b be as in the proof of Theorem 5.3. Note that the spaces L_a and L_b are preserved by the automorphism of R given by $f(X, Y) \mapsto f(X, -Y - a_1 X - a_3)$. Consider the $\mathbf{F}_q$-linear map

$$\vartheta' : L_a^q L_b \ \longrightarrow L_a L_b^q$$

defined by

$$e_i^q e_j \ \mapsto \ \overline{e_i} e_j^q.$$

Every function $F \in \ker \vartheta'$ vanishes on the set W. Indeed, let $F = \sum \lambda_{ij} e_i^q e_j$ for certain $\lambda_{ij} \in \mathbf{F}_q$ and let $P \in W$.

$$F(P)^q = \sum \lambda_{ij} e_i^{q^2}(P) e_j^q(P) = \sum \lambda_{ij} \overline{e}_i(P) e_j^q(P) = (\sum \lambda_{ij} \overline{e}_i f_j^q)(P) = \vartheta'(F)(P) = 0,$$

and hence $F(P) = 0$. Therefore we can draw the same conclusion as in the previous section. We have

$$\#\Omega^- \leq q^2 + 3q.$$

and hence

$$\begin{aligned} \#E(\mathbf{F}_{q^2}) - 1 &= \#\Omega^+, \\ &= \#\Omega - \#\Omega^- + \#(\Omega + \cap \#\Omega^-), \\ &\geq (2q^2 - r) - (q^2 + 3q) + r, \\ &\geq q^2 - 3q. \end{aligned}$$

as required.

Theorem 6.2. *(Hasse) The complex zeroes π and π' of the polynomial $T^2 - (h-q)T + q$ have absolute value $\sqrt{q}$. In particular $\pi' = \overline{\pi}$.*

Proof. The inequalities of section 5 and Lemma 6.4 provide us with the inequalities

$$q^d - 3q^{d/2} \leq q^d + 1 - \pi^d - \pi'^d \leq q^d + 3q^{d/2}, \qquad \text{for even } d \geq 0.$$

Therefore we have

$$|\pi^d + \pi'^d| \leq 3q^{d/2}, \qquad \text{for even } d \geq 0.$$

Suppose $|\pi| > \sqrt{q}$. Since $\pi\pi' = q$, we have $|\pi'| < \sqrt{q}$. Then the absolute values of both $1 + (\pi'/\pi)^d$ and $(\pi'/\pi)^d$ go to zero as $d \to \infty$. This is impossible. Therefore we have $|\pi| \leq \sqrt{q}$. By symmetry also $|\pi'| \leq \sqrt{q}$. This implies $|\pi| = |\pi'| = \sqrt{q}$, as required.

The inequalities of Theorem 3.4 and Proposition 4.1 have only been proved for $q \geq 5$. However, when $q < 5$, we have $q^d > 5$ when d is sufficiently large. This implies that we still have the inequality for large even degrees $d \geq 6$. Therefore the argument involving $d \to \infty$ is not affected and the conclusion is the same for $q < 5$. This proves the theorem.