The Riemann Hypothesis for Curves over Finite Fields

Aleksa Milojević

a senior thesis submitted in partial fulfillment
of the requirements for the degree of
Bachelor of Arts in Mathematics at
Princeton University

Advisor: Peter Sarnak

May 1, 2023

Contents

Acknowledgement

First of all, I would like to thank my advisor, Professor Peter Sarnak, for suggesting this topic and guiding me through it by suggesting many useful references. Without the numerous useful discussions and comments he provided, this work would hardly be possible. I would also like to thank Professor Zeev Dvir for accepting to be my second reader and for being one of the first people who introduced me to full power the polynomial method, which is the central topic of this thesis.

Further, I would like to thank Professor Noga Alon for advising my junior paper and for introducing me to the beautiful world of combinatorics. Prof. Alon, together all other professors of the Princeton Math Department, made my mathematical journey through Princeton a very special one.

Next, I thank Hans and Frank for helping me work through various proofs in this thesis. Finally, I must thank my friends Adam, Alex, Annabelle, Bingjian, Dafna, Danxian, Igor, Kiril, Lucas, and Marko for making my time at Princeton truly memorable.

I owe special gratitude to Jelisaveta, Marija and my parents, for their unceasing support and love.

Declaration

I declare that I have not violated the Honor Code during the composition of this work. This paper represents my own work in accordance with University regulations.

I authorize Princeton University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purposes of scholarly research.

Chapter 1

Introduction

The topic of this thesis is the zeta function of a curve over a finite field. Our goal will be to give a relatively elementary and self-contained proof of Artin's conjectures, using no heavy machinery from algebraic geometry besides the Riemann-Roch theorem. We try to keep the presentation as concrete as possible, by minimizing the amount of algebraic machinery used. Therefore, we present the proof only in the case of plane curves, which are sets of solutions to equations $f(x, y) = 0$. Of course, once a more elaborate algebraic setup has been performed, these methods apply almost unchanged to show the Riemann Hypothesis for general algebraic curves, i.e. for one-dimensional solution sets of systems of polynomial equations.

After introducing the main objects of interest in this chapter, Chapter 2 defines the basic elements of the theory of algebraic curves needed to show Artin's conjectures. Then, in Chapter 3, we will derive the Riemann-Roch theorem following the treatment of Niederreiter and Xing [22]. The Riemann-Roch theorem is one of the fundamental tools for investigating algebraic curves and we will use it to show the first two Artin conjectures, the rationality and functional equation for the zeta function, in Chapter 4. Then, in Chapter 5 we present a completely elementary point counting argument which will allow us to prove the Riemann Hypothesis for curves. To derive it, we will use a version of the polynomial method, known as the Stepanov method, and we follow the presentation given by Schmidt in [25]. This method will give us a slightly cruder bound than the one given by the Riemann Hypothesis, but we will be able to use the rationality and the functional equation of the zeta function to bootstrap it and obtain the full Riemann Hypothesis. Then, we complete the story by outlining another approach to proving the Riemann Hypothesis for curves, given by Bombieri in [4], also using a version of Stepanov's method.

However, the polynomial method was not the approach originally used to prove Riemann Hypothesis for curves - in the 1940s, Weil was the first one to prove the Riemann Hypothesis for curves, which use considerably more machinery from algebraic geometry. For more historical details an interested reader should consult Section 1.6 of this chapter.

1.1 Motivation: counting rational points on curves

One of the basic questions in number theory is trying to solve, or at least describe, rational solutions to various Diophantine equations and perhaps the most common variant of these equations are polynomial equations of the type $f(X_1, \ldots, X_n) = 0$. The theorems such as Hasse-Minkowski local-global principle show that the existence of rational solutions is often tightly linked to the existence of solutions modulo various primes. Hence, a very natural question arises: when does the equation $f(X_1, \ldots, X_n) \equiv 0 \pmod{p}$ have a solution? Furthermore, if such an equation has a solution, how many solutions are there?

In this thesis, we will focus on a special case when only two variables are present - we will try to understand the number of solutions to the equation $f(x, y) = 0$ modulo a prime number p. Amazingly, under some mild assumptions on the polynomial f, this number always turns out to be approximately p, with an error of order $O(p^{1/2})$. For example, the Hasse bound states that modulo any p, the number of solutions N to the equation $X^3 - X - Y^2 = 0$ in ${\mathbb{F}_p}^2$ always satisfies $|N - p| \leq 2\sqrt{p}$. More generally, we will consider the number of solutions

of $f(X,Y)=0$ in the field extensions $\mathbb{F}_p \subseteq \mathbb{F}_{p^2} \subseteq \cdots$ and we will show that this sequence of numbers is quite structured. In the case of the previously mentioned equation, $X^3 - X - Y^2 = 0$, if N_k denotes the number of solutions in field $\mathbb{F}_{p^k}$, we have $N_k = p^k + 1 - 2\mathrm{Re}\{\alpha^k\}$ for all $k \geq 1$, where α is a complex number of absolute value $|\alpha| = p^{1/2}$.

Let us now comment on the conditions our polynomials need to satisfy in order to be able to describe their solutions precisely. For example, if the polynomial $f(X,Y)$ factors as $f(X,Y) = f_1(X,Y)\cdots f_m(X,Y)$, then in order to count the solutions to $f(X,Y)=0$ precisely, one must understand not only how many solutions the equations $f_i(X,Y)=0$ have, but also how these solutions overlap. On one hand, the number of common solutions to different equations $f_i(X,Y)=0$ and $f_j(X,Y)=0$ can be bounded by $\deg f_i \cdot \deg f_j$. This means that the total number of solutions to $f(X,Y)=0$ is simply the sum of the corresponding numbers of solutions to equations $f_i(X,Y)=0$, up to a $O(m^2(\deg f)^2)$ error term. On the other hand, counting the number of solutions to $f(X,Y)=0$ exactly might be very challenging, since it amounts to counting the solutions to various systems of polynomial equations. Therefore, we will require the polynomial $f(X,Y)$ defining our equation to be irreducible throughout this thesis.

Finally, note that we are intending to consider the solutions to $f(X,Y)=0$ over all field extensions $\mathbb{F}_{p^k}$. Therefore, for the reasons described in the previous paragraph, it will be useful to assume that $f(X,Y)$ remains irreducible even after passing to a field extension of $\mathbb{F}_q$. In fact, it is not hard to see that this is equivalent to $f(X,Y)$ being irreducible over $\overline{\mathbb{F}_q}$. Hence, we make the following definition.

Definition 1.1. A polynomial $f \in \mathbb{F}_q[X,Y]$ is *absolutely irreducible* if it cannot be factored in the algebraic closure of $\mathbb{F}_q$, i.e. it cannot be written as a product $f = g \cdot h$, where $g, h \in \overline{\mathbb{F}_q}[X,Y]$ and $\deg g, \deg h < \deg f$.

Finally, as it is often done in algebraic geometry, we will sometimes work in projective space rather than in affine space. This allows us to count the intersections between various sets of solutions more easily and make the whole theory more symmetric. However, working in affine space also has its advantages, as it is often easier to define local properties in affine space and carry them over to projective space. Hence, throughout the thesis, we will also occasionally switch between the two. In fact, as we will see in Chapter 2, we can easily switch between affine and projective spaces while maintaining all relevant local properties of our curves.

1.2 Plane curves and their zeta functions

Let us begin by defining the main object we are interested in, the plane curves, which represent the set of solutions to polynomial equations in 2 variables.

Definition 1.2. Given a polynomial $f \in \overline{\mathbb{F}_q}[X,Y]$, the *affine plane curve* C defined by f is the set of points $(x,y) \in \overline{\mathbb{F}_q}^2$ for which $f(x,y)=0$. In the projective case, the definition is similar: for a homogeneous polynomial $F \in \overline{\mathbb{F}_q}[X,Y,Z]$, the *projective plane curve* C defined by F is the set of points $[x:y:z] \in \mathbb{P}^2(\overline{\mathbb{F}_q})$ for which $F(x,y,z)=0$.[1]

If a polynomial f defining a plane curve C lies in $\mathbb{F}_q[X,Y]$, we can also consider C as a curve over the field $\mathbb{F}_q$ and define the $\mathbb{F}_q$-rational points of C the points whose coordinates lie in $\mathbb{F}_q$. Similarly, for a projective plane curve C, we define $\mathbb{F}_q$-rational points of C as the set of points which can be written as $[x:y:z]$ for some $x,y,z \in \mathbb{F}_q$, not all zero. Apart from being irreducible, for certain parts of the argument our curves will also need to be smooth. This is a slightly stronger condition than irreducibility, which intuitively ensures that the curve has no "cusps" and does not "cross itself".

Definition 1.3. If $f \in \overline{\mathbb{F}_q}[X,Y]$ is the defining equation of the affine plane curve C, we say that C is *smooth* if no point $(x,y) \in C$ satisfies the system of equations $\partial_X f(x,y) = \partial_Y f(x,y) = 0$. Similarly, given a homogeneous polynomial $F \in \overline{\mathbb{F}_q}[X,Y,Z]$ and the projective plane curve C defined by F, we say that C is *smooth* if no point $[x:y:z] \in C$ satisfies the system of equations $\partial_X F(x,y,z) = \partial_Y F(x,y,z) = \partial_Z F(x,y,z) = 0$.

For a smooth projective curve, we are ready to introduce the central object of this thesis.

[1] For a formal definition of projective space $\mathbb{P}^2(\overline{\mathbb{F}_q})$, the reader should consult Chapter 2.

Definition 1.4. The *zeta function* associated to the projective curve $C/\mathbb{F}_q$ defined by an absolutely irreducible homogeneous polynomial $F \in \mathbb{F}_q[X, Y, Z]$ is defined as

$$Z(C/\mathbb{F}_q, T) = \exp\left(\sum_{m\geq 1} \frac{\#C/\mathbb{F}_{q^m}}{m} T^m\right),$$

where $\#C/\mathbb{F}_{q^m}$ is the number of $\mathbb{F}_{q^m}$-rational points on the curve C.

Remark 1.5. Although this definition may seem unmotivated, since there are many possible ways to define a generating function from the sequence of numbers $(\#C/\mathbb{F}_{q^m})_{m\geq 1}$, we will see in a later section that, after a change of coordinates, this definition becomes very much analogous to the Riemann zeta function defined over the complex numbers.

On the other hand, once we introduce the notion of a divisor, we will show that $Z(C/\mathbb{F}_q, T) = \sum_{d\geq 0} D_d T^d$, where D_d is the number of nonnegative divisors of degree d. Hence, from the perspective of divisors, the zeta function as defined above is a very natural object to consider.

Now we are ready to state the main theorem shown in this thesis, which was first conjectured by Artin. The first two assertions of this theorem were proven by F.K. Schmidt and the third assertion was shown by Weil.

Theorem 1.6. Let $F \in \mathbb{F}_q[X, Y, Z]$ be an absolutely irreducible homogeneous polynomial and let $C/\mathbb{F}_q$ be the corresponding projective plane curve, which we assume to be smooth. Then, there exists a positive integer g, depending on F such that the zeta function of the curve $C/\mathbb{F}_q$ has the following properties:

- $Z(C/\mathbb{F}_q, T)$ is a rational function of T. Even more precisely, there exists a polynomial $L(T)$ with integer coefficients of degree $2g$ such that

$$Z(C/\mathbb{F}_q, T) = \frac{L(T)}{(1-T)(1-qT)}, \tag{1.1}$$

- $Z(C/\mathbb{F}_q, T)$ satisfies the following functional equation

$$Z(C/\mathbb{F}_q, T) = T^{2g-2} q^{g-1} Z\left(C/\mathbb{F}_q, \frac{1}{qT}\right), \tag{1.2}$$

- Finally, all roots of L are complex numbers of absolute value $q^{-1/2}$.

Remark 1.7. The integer g mentioned in the above theorem is called the *genus* of the curve, and represents one of the fundamental quantities related to the curve.

Remark 1.8. The third statement of the above theorem is called the *Riemann Hypothesis for curves*, owing to the connection to the usual Riemann Hypothesis which will be presented in the next section.

Let us describe, in a concrete way, what Theorem 1.6 can tell us about counting the number of points on C over $\mathbb{F}_{q^m}$. We begin by an interpretation of the functional equation in terms of the roots of the polynomial $L(T)$ and then show a formula for $\#C/\mathbb{F}_{q^m}$ based on these roots.

Before we state the first corollary of Theorem 1.6, let us just note that $L(0) = Z(C/\mathbb{F}_q, 0) = 1$ and therefore one can write $L(T)$ as $L(T) = \prod_{i=1}^{2g}(1 - \alpha_i T)$, for some $\alpha_1, \ldots, \alpha_{2g} \in \mathbb{C}$.

Corollary 1.9. Suppose that $Z(C/\mathbb{F}_q, T)$ is rational, satisfying the formula (1.1), and satisfies the functional equation (1.2). If we write the polynomial $L(T)$ in the form $L(T) = \prod_{i=1}^{2g}(1 - \alpha_i T)$, the values α_i satisfy, after a possibly permuting them, $\alpha_i \alpha_{2g+1-i} = q$ for all $i = 1, \ldots, g$.

Proof. Expanding out the functional equation (1.2) using the formula (1.1), we find $L(T) = q^g t^{2g} L\left(\frac{1}{qT}\right)$. In terms of the values $\alpha_1, \ldots, \alpha_{2g}$, this equation can be expressed as

$$\prod_{i=1}^{2g}(1 - \alpha_i T) = q^g T^{2g} \prod_{i=1}^{2g}\left(1 - \frac{\alpha_i}{Tq}\right).$$

Comparing the coefficients of these polynomials, we find that $\prod_{i=1}^{2g} \alpha_i = q^g$. Hence, we can rearrange the above expression with the goal of comparing the roots of left and right hand sides.

$$\prod_{i=1}^{2g}(1-\alpha_i T) = q^g \prod_{i=1}^{2g}\left(T - \frac{\alpha_i}{q}\right) = q^{2g}\prod_{i=1}^{2g}\left(T\alpha_i^{-1} - \frac{1}{q}\right) = \prod_{i=1}^{2g}\left(1 - T\frac{q}{\alpha_i}\right).$$

Hence, the map $\alpha_i \mapsto q/\alpha_i$ permutes the set $\{\alpha_1, \ldots, \alpha_{2g}\}$ and therefore we may order $\alpha_1, \ldots, \alpha_{2g}$ such that $\alpha_i \alpha_{2g+1-i} = q$ for all $i = 1, \ldots, g$. □

Corollary 1.10. Under the assumptions of Corollary 1.9, we have the following formula for the number of points on C over $\mathbb{F}_{q^m}$:

$$\#C/\mathbb{F}_{q^m} = q^m + 1 - \sum_{i=1}^{2g} \alpha_i^m. \tag{1.3}$$

Proof. The main idea is to take formal logarithm of the equation (1.1) and expand the polynomials on the right hand side using Taylor expansion. More precisely, we have:

$$\begin{aligned}\sum_{m\geq 1} \#C/\mathbb{F}_{q^m}\frac{T^m}{m} = \log Z(C/\mathbb{F}_q, T) &= \sum_{i=1}^{2g}\log(1-\alpha_i T) - \log(1-T) - \log(1-qT)\\ &= \sum_{m\geq 1}\frac{T^m}{m} + \frac{(qT)^m}{m} - \sum_{i=1}^{2g}\frac{(T\alpha_i)^m}{m}\\ &= \sum_{m\geq 1}\frac{T^m}{m}\left(1 + q^m - \sum_{i=1}^{2g}\alpha_i^m\right).\end{aligned}$$

The formula (1.3) is now derived by comparing the coefficients next to T^m. □

Note that Corollaries 1.9 and 1.10 used only the rationality and functional equation for the zeta function, which are comparatively easy to show. However, combining the Riemann Hypothesis for curves with the formula (1.3) gives a precise estimate on the number of $\mathbb{F}_{q^m}$-rational points of C, which is very reminiscent to the previously mentioned Hasse bound.

Corollary 1.11. If Theorem 1.6 holds, we have

$$|\#C/\mathbb{F}_{q^m} - (q^m + 1)| \leq 2gq^{m/2}. \tag{1.4}$$

Proof. The proof is immediate, since the Riemann Hypothesis for curves implies that $|\alpha_i| = q^{1/2}$. Combined with the formula (1.3), this implies the desired bound via the triangle inequality. □

The bound (1.4) is tight if q is a even power of a prime. The curves attaining the bound are called *maximal* or *minimal*, depending on whether the number of points of them attains the upper or the lower bound posed by the bound (1.4).

If we assume the rationality and the functional equation for the zeta function, we can also show the reverse implication to Corollary 1.11. More precisely, if we have $|\#C/\mathbb{F}_{q^m} - (q^m + 1)| \leq K_{C/\mathbb{F}_q} q^{m/2}$ for some constant $K_{C/\mathbb{F}_q}$ depending only on the curve C, we can derive $|\alpha_i| = q^{1/2}$ for all i. This will be very important since the proofs of the Riemann Hypothesis for curves will rely on establishing the bound on the number of points of $C/\mathbb{F}_{q^m}$.

Proposition 1.12. Suppose that the curve $C/\mathbb{F}_q$ has a rational zeta function, satisfying the functional equation (1.2). If there exists a constant K (possibly depending on the curve $C/\mathbb{F}_q$) and an integer m for which

$$|\#C/\mathbb{F}_{q^{mn}} - (q^{mn} + 1)| \leq Kq^{mn/2}, \tag{1.5}$$

for all $n \geq 1$, then the Riemann Hypothesis holds for the curve $C/\mathbb{F}_q$ (in other words, $|\alpha_i| = q^{1/2}$ for all i).

Proof. The starting point of this proof is the trace formula (1.3), which states that $q^{mn} + 1 - \#C/\mathbb{F}_{q^{mn}} = \sum_{i=1}^{2g} \alpha_i^{mn}$. Moreover, assume that $\alpha_1, \ldots, \alpha_{2g}$ are arranged in decreasing order of absolute value.

First, we will use Dirichlet's approximation theorem to show that for there are infinitely many values of n such that for all i we have $\mathrm{Re}\left(\frac{\alpha_i}{\alpha_1}\right)^{mn} \geq 0$. In other words, we have to ensure that $mn \arg(\alpha_i/\alpha_1) \bmod \pi \in [-\pi/2, \pi/2]$. But the simultaneous version of Dirichlet theorem states that for any N, there exists $n \leq N$ such that $mn \arg(\alpha_i/\alpha_1) \bmod \pi \in [-\pi N^{1/d}, \pi N^{1/d}]$. This suffices to construct an infinite sequence of n satisfying the above property.

For n chosen above, we have

$$|\sum_{i=1}^{2g} \alpha_i^{mn}| = |\alpha_1|^{mn} \left|\sum_{i=1}^{2g} \left(\frac{\alpha_i}{\alpha_1}\right)^{mn}\right| \geq |\alpha_1|^{mn}.$$

Hence, we have $|\alpha_1| \leq K^{1/mn} q^{1/2}$ and as $n \to \infty$ we obtain $|\alpha_1| \leq q^{1/2}$. By our choice of ordering of $\alpha_1, \ldots, \alpha_{2g}$ we obtain $|\alpha_i| \leq q^{1/2}$ for all i. But Corollary 1.9 requires that $\alpha_i \alpha_{2g+1-i} = q$ (possibly under a different ordering), which means that we must have equality and so $|\alpha_i| = q^{1/2}$ for all i. □

1.3 Divisors on a curve

In this section, we define the notion of a divisor on the curve, which we will use repeatedly throughout this thesis. For a curve over an algebraically closed field $\overline{\mathbb{F}_q}$, this notion is simple enough to define. A *divisor* on a curve $C/\overline{\mathbb{F}_q}$ is a formal expression of the form $\sum_{P \in C} n_P \cdot P$, where n_P are integers and we have $n_P = 0$ for all but finitely many points $P \in C$. One can then define the operation of addition on divisors, by simply adding the corresponding coefficients n_P, making them into an abelian group. In a more abstract language, a divisor is an element of the free abelian group on the points of the curve.

However, this definition is not descriptive enough for curves over non-algebraically closed fields. The philosophy we adopt here is that the group of divisors on the curve should contain the information about the points of the curve over the base field itself as well as over all algebraic extension of this field. Let us illustrate this by a concrete example. The sets of points defined by the equations $X + Y + Z = 0$ and $X^3 + Y^3 + Z^3 = 0$ are precisely the same over $\mathbb{F}_2$, but these curves are definitely not the same over $\overline{\mathbb{F}}_2$. Hence, if we considered the set of divisors only as formal sums of points on the curve itself, we would not be able to distinguish between the above two curves by the means of divisors, let alone count the number of points in the field extensions. Hence, we adopt a slightly more intricate definition of a divisors, which relies on the action of the Galois group on the $\overline{\mathbb{F}_q}$-points of the curve.

The Galois group $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$ is the group of automorphisms of $\overline{\mathbb{F}_q}$ which fix $\mathbb{F}_q$ and it is an abelian group generated by the Frobenius automorphism $\sigma : x \mapsto x^q$ for $x \in \overline{\mathbb{F}_q}$. This group acts on the points $[x : y : z] \in C$ by simply acting on their coordinates, meaning that we define $\sigma([x : y : z]) = [\sigma(x) : \sigma(y) : \sigma(z)]$ for $\sigma \in Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$.

Definition 1.13. A *prime divisor* P on a curve $C/\mathbb{F}_q$ is a $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$-orbit of points on $C/\overline{\mathbb{F}_q}$. A *divisor* D on the curve $C/\mathbb{F}_q$ is a formal linear combination $\sum_P n_P \cdot P$ of prime divisors with integer coefficients, where all but finitely many coefficients n_P are zero. Finally, we say that a divisor is *effective* if all of its coefficients are nonnegative.

This definition implicitly assumes that the curve $C/\overline{\mathbb{F}_q}$ is stable under the action of the Galois group $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$. In other words, we assume that if a point $[x : y : z] \in C/\overline{\mathbb{F}_q}$, then $[\sigma(x) : \sigma(y) : \sigma(z)] \in C/\overline{\mathbb{F}_q}$ too. This is not hard to see, since if $F \in \mathbb{F}_q[X, Y, Z]$ is the defining polynomial of C we have $f(\sigma(x), \sigma(y), \sigma(z)) = f(x, y, z) = 0$ for all $[x : y : z] \in C/\mathbb{F}_q$.

Let us note that prime divisors are sometimes referred to as *places* in the literature. The term place comes from algebraic number theory, where a place of a number field roughly referring to a valuation on the field. For the further discussion of the analogy between prime divisors and places of the function fields of the curve, consult Sections 1.4, 2.2 and 2.3.

As before, the divisors on a curve form an abelian group under the operation of addition. The group of

divisors on a curve $C/\mathbb{F}_q$ will be denoted by $\mathrm{Div}(C/\mathbb{F}_q)$, and the set of prime divisors of $C/\mathbb{F}_q$ will be denoted by $\mathrm{PDiv}(C/\mathbb{F}_q)$.[2]

Finally, the above definition of a divisor comes with a word of warning. Namely, as we defined them, the divisors on the curve $C/\mathbb{F}_q$ depend on the base field $\mathbb{F}_q$, and hence the set of prime divisors of $C/\mathbb{F}_q$ and $C/\mathbb{F}_{q^m}$ will not be the same in general.

Definition 1.14. The *degree* of a prime divisor P is the number of elements in the $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$-orbit defining this divisor, which we denote by $\deg P$ or simply $|P|$. We extend this definition linearly and define the degree of a general divisor $D = \sum_P n_P \cdot P$ as $\deg D = \sum_P n_P \cdot \deg P$, where the sum runs over all prime divisors of a curve.

A priori, it is not obvious that a degree of a prime divisor is finite at all. However, we have the following simple proposition shows this.

Proposition 1.15. If P is a prime divisor on a curve $C/\mathbb{F}_q$, then P is a finite set.

Proof. In simpler terms, this proposition states that a $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$-orbit of any point of $[x : y : z] \in C/\overline{\mathbb{F}_q}$ is finite. Since $x, y, z \in \overline{\mathbb{F}_q} = \bigcup_{m=1}^{\infty} \mathbb{F}_{q^m}$, there exists a finite extension $\mathbb{F}_{q^m}$ of $\mathbb{F}_q$ such that all of x, y, z are in $\mathbb{F}_{q^m}$. Then, the action of any element of $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$ leaves x, y and z within $\mathbb{F}_{q^m}$, meaning that P is a subset of $\mathbb{P}^2_{\mathbb{F}_{q^m}}$, and thus finite. □

As indicated above, the group of divisors of the curve will be the key object in understanding points of the curve $C/\mathbb{F}_{q^m}$ over various algebraic extensions of $\mathbb{F}_q$. Let us illustrate this by an example before showing the general theorem.

Example 1.16. Consider the Fermat curve C given by the equation $X^3 + Y^3 + Z^3 = 0$ over $\mathbb{F}_2$. The prime divisors of this curve of degree 1 correspond to those points of $C/\overline{\mathbb{F}}_2$ fixed by the Galois group $Gal(\overline{\mathbb{F}}_2/\mathbb{F}_2)$, which are precisely the points whose coordinates are in $\mathbb{F}_2$. There are three such points on C: $[1 : 1 : 0], [1 : 0 : 1]$ and $[0 : 1 : 1]$. Hence, C has three prime divisors of degree 1.

How do we find prime divisors of degree 2 on this curve? Just a little Galois theory is all we need. Namely, if P is a prime divisor of degree 2, i.e. just a two-element orbit $\{[x : y : z], [x' : y' : z']\}$ of points on $C/\overline{\mathbb{F}}_2$, we will show that the coordinates of these points must lie in $\mathbb{F}_4$. Let $\mathbb{F}_{2^m}$ be the smallest extension of $\mathbb{F}_2$ containing x. Then, the $Gal(\overline{\mathbb{F}}_2/\mathbb{F}_2)$-orbit of x is the same as $Gal(\mathbb{F}_{2^m}/\mathbb{F}_2)$ orbit of x. Furthermore, we know that group $Gal(\mathbb{F}_{2^m}/\mathbb{F}_2)$ is cyclic group generated by the automorphism $\sigma : \mathbb{F}_{2^m} \to \mathbb{F}_2$ given by $\sigma(\alpha) = \alpha^2$. Since the orbit of x has only two elements, we conclude that $\sigma^2(x) = x$, i.e. $x^4 - x = 0$. Hence, $x \in \mathbb{F}_4$, as we have claimed.

The search for the prime divisors of degree 2 is now simplified considerably - it suffices to consider only the points of $C/\mathbb{F}_4$. To characterize the points of $C/\mathbb{F}_4$, note the following elementary statement: for $\alpha \in \mathbb{F}_4$, we have $\alpha^3 = 1$ if and only if $\alpha \neq 0$. Hence, if $[x : y : z] \in C/\mathbb{F}_4$, exactly one of the coordinates x, y, z must be zero, which gives the following set of points on $C/\mathbb{F}_4$ (where we denote the elements of $\mathbb{F}_4$ by $0, 1, \alpha, \alpha + 1$ with $\alpha^2 = \alpha + 1$):

$$C/\mathbb{F}_4 = \{[1 : 1 : 0], [1 : \alpha : 0], [1 : \alpha + 1 : 0], [1 : 0 : 1], [\alpha : 0 : 1], [\alpha + 1 : 0 : 1], [0 : 1 : 1], [0 : 1 : \alpha], [0 : 1 : \alpha + 1]\}.$$

Note that the points with coordinates in $\mathbb{F}_2$ are prime divisors of degree 1, as discussed above. Hence, the prime divisors of degree 2 are the pairs $\{[1 : \alpha : 0], [1 : \alpha + 1 : 0]\}, \{[0 : 1 : \alpha], [0 : 1 : \alpha + 1]\}, \{[\alpha : 0 : 1], [\alpha + 1 : 0 : 1]\}$. We could continue looking for divisors of higher degrees in a similar way, but enumerating all points of $C/\mathbb{F}_{2^m}$ would soon be out of reach. Besides, we introduced divisors to help us count points on curves, and hence we must find a better way to enumerate prime divisors on curves without enumerating the points of the curve first.

Now, we will express the relationship between divisors and the rational points on the curve in a more general setting.

Proposition 1.17. Let $P = \{P_1, \ldots, P_d\}$ be a prime divisor of degree d on a curve $C/\mathbb{F}_q$. Then P is partitioned into (m, d) prime divisors of $C/\mathbb{F}_{q^m}$, and each of these prime divisors has degree $\frac{d}{(m,d)}$. In particular, if $d|m$, all points of P are $\mathbb{F}_{q^m}$-rational.

[2] This notation should not be confused with the set of principal divisors, which will be introduced later.

Proof. This proof is an exercise in Galois theory of finite fields. By definition, $\{P_1, \dots, P_d\}$ is a $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$-orbit of point P_1, and hence we may relabel the points such that $P_i = \sigma^{i-1}(P_1)$, where $\sigma : [x : y : z] \mapsto [x^q : y^q : z^q]$ denotes the Frobenius automorphism, and $\sigma^d(P_1) = P_1$. Further, we know that prime divisors of $C/\mathbb{F}_{q^m}$ correspond to $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_{q^m})$-orbits of points, and that $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_{q^m})$ is generated by $\sigma^m : [x : y : z] \mapsto [x^{q^m} : y^{q^m} : z^{q^m}]$. Hence, the prime divisors of $C/\mathbb{F}_{q^m}$ arising from P are $\{P_1, \sigma^m(P_1), \dots\}$, $\{P_2, \sigma^m(P_2), \dots\}$ etc. Elementary modular arithmetic tells us that this splits the set $\{P_1, \sigma(P_1), \dots, \sigma^{d-1}(P_1)\}$ into $\frac{d}{(m,d)}$ sets of cardinality (m, d). □

Corollary 1.18. Let $C/\mathbb{F}_q$ be a smooth curve, and let $\#C/\mathbb{F}_{q^m}$ denote the number of points on $C/\mathbb{F}_{q^m}$. Then, we have

$$\#C/\mathbb{F}_{q^m} = \sum_{P \in \mathrm{PDiv}(C/\mathbb{F}_q), \deg P | m} \deg P.$$

Proof. The proof relies on the application of the special case of Proposition 1.17. Namely, the rational points of $C/\mathbb{F}_{q^m}$ are contained in the prime divisors P of $C/\mathbb{F}_q$ with $\deg P | m$, and every such prime divisor contains exactly $(m, \deg P) = \deg P$ such points. The statement then follows directly. □

1.4 Relation to Riemann zeta function

In this section, we will show the analogy between the zeta functions associated to algebraic curves and Riemann zeta function $\zeta(s)$. Although this section is not required for the proof of the Theorem 1.6, it provides useful context and motivation Artin conjectures. Parts of this discussion will therefore assume familiarity with notions in algebraic number theory.

Let us begin by introducing the Riemann zeta function. It is defined by the equation $\zeta(s) = \sum_{n \geq 1} n^{-s}$ in the region $\Re s > 1$ can be extended to a meromorphic function on the whole complex plane. Moreover, it possesses a number of properties relating it to the arithmetic structure of the integers, such as the Euler product

$$\zeta(s) = \prod_{p \text{ prime}} (1 - p^{-s})^{-1}.$$

This notion can be extended to arbitrary an number field K by defining

$$\zeta_K(s) = \prod_{\mathfrak{p} \subseteq \mathcal{O}_K} (1 - N_{K/\mathbb{Q}}(\mathfrak{p})^{-s})^{-1},$$

where the product ranges over all prime ideals $\mathfrak{p}$ within the ring of integers $\mathcal{O}_K$ and $N_{K/\mathbb{Q}}$ denotes the ideal norm.

For curves, one might try to follow a similar recipe and define the zeta function

$$\zeta(C/\mathbb{F}_q, s) = \prod_{P \in \mathrm{PDiv}(C/\mathbb{F}_q)} (1 - q^{-\deg P \cdot s})^{-1},$$

where $q^{\deg P}$ plays the role of the norm. One might complain that the zeta function we just defined has nothing to do with $Z(C/\mathbb{F}_q, T)$ from definition 1.4. Despite apparent differences, under the changes of coordinates $q^{-s} = T$ these two functions become exactly the same.

Proposition 1.19. The zeta function of the curve $C/\mathbb{F}_q$, defined in Definition 1.4, can be written as

$$Z(C/\mathbb{F}_q, T) = \prod_{P \in \mathrm{PDiv}(C/\mathbb{F}_q)} \frac{1}{1 - T^{\deg P}} = \sum_{D \in \mathrm{Div}(C/\mathbb{F}_q), D \geq 0} T^{\deg D}. \tag{1.6}$$

Hence, we have the equality $Z(C/\mathbb{F}_q, T) = \zeta(C/\mathbb{F}_q, s)$ when $T = q^{-s}$.

Proof. This proof is based the relation between divisors and rational points of a curve presented in Corollary 1.18. If we denote the number of prime divisors of degree d by P_d, the proof reduces to the following straightforward

manipulation of the formal power series.

$$Z(C/\mathbb{F}_q,t)=\exp\left(\sum_{m\geq 1}\frac{\#C/\mathbb{F}_q}{m}t^m\right)=\exp\left(\sum_{m\geq 1}\frac{\sum_{d|m}d\cdot P_d}{m}t^m\right)=\exp\left(\sum_{d\geq 1}P_d\sum_{m=kd,k\geq 1}\frac{d}{m}t^m\right)=$$
$$=\exp\left(\sum_{d\geq 1}P_d\sum_{k\geq 1}\frac{1}{k}t^{kd}\right)=\exp\left(\sum_{P\in\mathrm{PDiv}(C/\mathbb{F}_q)}-\log(1-t^{\deg P})\right)=\prod_{P\in\mathrm{PDiv}(C/\mathbb{F}_q)}\frac{1}{1-t^{\deg P}}=$$
$$=\prod_{P\in\mathrm{PDiv}(C/\mathbb{F}_q)}\sum_{n_P\geq 0}t^{n_P\deg P}=\sum_{D\in\mathrm{Div}(C/\mathbb{F}_q),D\geq 0}t^{\deg D}.$$

□

Right now, we are able to see the similar form of the definitions of $\zeta(C/\mathbb{F}_q,s)$ and $\zeta(s)$. But the analogy between the two situations extends further. Namely, prime divisors of $C/\mathbb{F}_q$ induce valuations on the function field $\mathbb{F}_q(C)$ (for the formal discussion of this consult Sections 2.1 and 2.3). Even more is true - the places of $\mathbb{F}_q(C)$ are in one to one correspondence with prime divisors $P\in\mathrm{PDiv}(C/\mathbb{F}_q)$. This fact will not be used in our proofs, and hence we will not derive it. An interested reader should consult Section 3.1 of [22].

On the other hand, the non-Archimedean places of a number field K correspond exactly to the prime ideals of $\mathcal{O}_K$. Hence, we see that even from this perspective, the products appearing in $\zeta_K(s)$ and $\zeta(C/\mathbb{F}_q,s)$ are completely analogous.

Finally, the notion of the norm of the ideal $\mathfrak{p}\subset\mathcal{O}_K$ is usually defined by $N_{K/\mathbb{Q}}(\mathfrak{p})=|\mathcal{O}_K/\mathfrak{p}|$. In other words, the norm of the ideal is the size of its residue field. In Proposition 2.21, we will show that the size of the residue field of the place corresponding to $P\in\mathrm{PDiv}(C/\mathbb{F}_q)$ is precisely $q^{\deg P}$. Hence, we see that the natural replacement for $N_{K/\mathbb{Q}}(\mathfrak{p})$ is precisely $q^{\deg P}$, which then explains the analogy between the definitions of $\zeta_K(s)$ and $\zeta(C/\mathbb{F}_q,s)$.

1.5 Two examples

In this section, we will focus on a series of simple examples that illustrate the behavior of zeta functions described in the Theorem 1.6.

Example 1.20 (Linear polynomials.). Probably the simplest kind of equation one can try to solve over a finite field is a linear equation of the type $f(X,Y,Z)=aX+bY+cZ=0$ where at least one of a,b,c is nonzero, say $a\neq 0$. Then, it is not hard to see that we have exactly q^m+1 solutions in $\mathbb{P}^2_{\mathbb{F}_q}$. Namely, if $a\neq 0$, the set of solutions is precisely $\{[\frac{-by-cz}{a}:y:z]|[y:z]\in\mathbb{P}^1(\mathbb{F}_{q^m})\}$. Hence, the zeta function of the line L defined by f is

$$Z(L/\mathbb{F}_q,t)=\exp\left(\sum_{m\geq 1}\frac{\#C/\mathbb{F}_{q^m}}{m}t^m\right)=\exp\left(\sum_{m\geq 1}\frac{(qt)^m}{m}+\sum_{m\geq 1}\frac{t^m}{m}\right)$$
$$=\exp\left(-\log(1-qt)-\log(1-t)\right)=\frac{1}{(1-t)(1-qt)}.$$

Note that the solution set of the equation $aX+bY+cZ=0$ corresponds to the projective line $\mathbb{P}^1(\mathbb{F}_q)$, which has the same zeta functions as the one we just computed.

Example 1.21. In this example, we will consider the zeta function of the Fermat cubic $C/\mathbb{F}_q$ given by $X^3+Y^3+Z^3=0$. Assuming the rationality of the zeta function, we will verify the functional equation and demonstrate the Riemann Hypothesis for $q=2$. Then, we will show a general and completely elementary way to verify the bound (1.4).

The genus of the cubic $C/\mathbb{F}_q$ is $g=1$, and therefore the rationality of $Z(C/\mathbb{F}_q,T)$ implies it can be written as

$$Z(C/\mathbb{F}_q,T)=\frac{L(T)}{(1-T)(1-qT)},$$

for a degree 2 polynomial $L(T)$. Since $L(0)=Z(0)=1$, we can write $L(T)=(1-\alpha_1T)(1-\alpha_2T)$, where α_1,α_2 are complex numbers. One can determine α_1,α_2 from the trace formula (1.3), which says that $C/\mathbb{F}_{q^m}=$

$q^m + 1 - \alpha_1^m - \alpha_2^m$. Since we have only two unknowns, α_1 and α_2, it suffices to check the number of solutions in the fields $\mathbb{F}_2, \mathbb{F}_4$. We have already done this in Example 1.16, and we have found that $\#C/\mathbb{F}_2 = 3, \#C/\mathbb{F}_4 = 9$. Hence, we have $\alpha_1 + \alpha_2 = 0$ and $\alpha_1^2 + \alpha_2^2 = -4$, giving us the solutions $\alpha_1 = i\sqrt{2}, \alpha_2 = -i\sqrt{2}$ and the zeta function

$$Z(C/\mathbb{F}_2, T) = \frac{1 + 2T^2}{(1-T)(1-2T)}.$$

Now, we can easily verify the functional equation, since

$$Z(C/\mathbb{F}_2, \frac{1}{2T}) = \frac{1 + 2(1/2T)^2}{(1-1/2T)(1-1/T)} = \frac{2T^2+1}{(2T-1)(T-1)} = Z(C/\mathbb{F}_2, T).$$

Furthermore, note that the Riemann Hypothesis also holds because $|\alpha_1| = |\alpha_2| = \sqrt{2}$.

Of course, this holds for larger values of q as well. We will now present a theorem of Gauss which shows that for any $\mathbb{F}_q$, the number of $\mathbb{F}_q$-rational points of $C/\mathbb{F}_q$ satisfies

$$|\#C/\mathbb{F}_q - (q+1)| \leq 2q^{1/2}.$$

Note that this bound holds also for all power of q, by replacing the base field $\mathbb{F}_q$ by $\mathbb{F}_{q^m}$. As discussed in Corollary 1.11 and Proposition 1.12, this is equivalent to the Riemann Hypothesis for C (since the genus of C is $g = 1$). Let us now state Gauss' theorem.

Theorem 1.22. Consider the curve $C/\mathbb{F}_q$ given by $X^3 + Y^3 + Z^3 = 0$, where $q = p^k$ and $p \equiv 1 \pmod 3$. Then there exist integers A, B for which $4q = A^2 + 27B^2$ and $\#C/\mathbb{F}_q = q + 1 + A$.

Before we prove this theorem, let us note that since $4q \geq A^2$ we must have $|\#C/\mathbb{F}_q - (q+1)| = A \leq 2q^{1/2}$, which is equivalent to the Riemann Hypothesis for $C/\mathbb{F}_q$ as discussed above. Also, note that the assumption $p \equiv 1 \pmod 3$ can be replaced by the assumption that not all elements are cubes in $\mathbb{F}_q$. Of course, if all elements are indeed cubes, then the Fermat cubic has the same number of points as $X + Y + Z = 0$, which was already considered in the previous example.

Proof of Theorem 1.22. Our presentation follows the treatment of Silverman, Tate [27] and Mazzoni and Schiltknecht [19]. We will consider the set of nonzero cubes $R = \{x^3 : x \in \mathbb{F}_q\}\backslash\{0\}$ and for sets $A, B, C \subseteq \mathbb{F}_q$ we will define the symbol

$$[ABC] = |\{(a,b,c) \in A \times B \times C : a + b + c = 0\}|.$$

It is not hard to see that this symbol remains unchanged when A, B, C are permuted, or all scaled by the same nonzero constant $\lambda \in {\mathbb{F}_q}^\times$.

Let us begin by observing that we have exactly 9 solutions to $X^3 + Y^3 + Z^3 = 0$ when one of X, Y, Z is zero. This is because, when $Z = 0$, the solutions to $X^3 + Y^3 = 0$ can be explicitly found to be $\{[1 : \zeta : 0], [1 : \zeta^2 : 0], [1 : -1 : 0]\}$, where $\zeta \in \mathbb{F}_q$ is the root of $\zeta^2 + \zeta + 1 = 0$.

Now, note that the number of points on $C/\mathbb{F}_q$ with all coordinates nonzero is $\frac{27[RRR]}{p-1}$, since every triple $(r_1, r_2, r_3) \in R \times R \times R$ with $r_1 + r_2 + r_3 = 0$ yields 27 ordered triples (x, y, z) with $x^3 + y^3 + z^3 = 0$, since every r_i has three cube roots in $\mathbb{F}_q$. However, since we are working in projective space and we do not count scaled solutions as different, we divide by $q - 1$. Denoting $|R| = m = \frac{q-1}{3}$, we obtain the formula for $\#C/\mathbb{F}_q$ which reads

$$\#C/\mathbb{F}_q = 9 + \frac{9[RRR]}{m}.$$

Let us now define cosets of R in the multiplicative group ${\mathbb{F}_q}^\times$ to be S, T, with the goal of showing $\#C/\mathbb{F}_q = 9[RST]/m$. To do this, we note that $[RR\mathbb{F}_q] = m^2$ and therefore

$$[RR\{0\}] + [RRR] + [RRS] + [RRT] = m^2.$$

Similarly, we have $[ST\mathbb{F}_q] = m^2$ and therefore

$$[ST\{0\}] + [STR] + [STS] + [STT] = m^2.$$

Picking an arbitrary element $s \in S$ and noting that $sR = S, sS = T, sT = R$ gives $[RRS] = [SST]$ and $[TTS] = [RRT]$. Moreover, as discussed before we have $[RR\{0\}] = m$, since R is symmetric $R = -R$. Finally,

we have $[ST\{0\}] = 0$, since $S \cap (-T) = \emptyset$. This is easy to see since both S and T are closed under multiplications by cubes and therefore symmetric $s \in S \Longrightarrow (-1)^3 s = -s \in S$.

Combining these observations gives $[RRR] + m = [RRR] + [RR\{0\}] = [RTS]$ and so

$$\#C/\mathbb{F}_q = 9 + \frac{9[RTS]}{m}.$$

Let $\omega_1, \ldots, \omega_k$ be primitive complex p-th roots of unity. We know that we can regard $\mathbb{F}_q$ as a k-dimensional vector space over $\mathbb{F}_p$. Let us fix an arbitrary basis and write every $x \in \mathbb{F}_q$ as $x = (x_1, \ldots, x_k) \in {\mathbb{F}_p}^k$. Then, we define $\alpha_1 = \sum_{x \in R} \omega^x, \alpha_2 = \sum_{x \in S} \omega^x, \alpha_3 = \sum_{x \in T} \omega^x$, where ω^x is a shorthand notation for $\omega_1^{x_1} \cdots \omega_k^{x_k}$. Then, the complex numbers $\alpha_1, \alpha_2, \alpha_3$ satisfy the following properties

- $\alpha_1 + \alpha_2 + \alpha_3 = -1$,
- $\alpha_1\alpha_2 + \alpha_2\alpha_3 + \alpha_3\alpha_1 = -m$,
- $\alpha_1\alpha_2\alpha_3 = am - \frac{m^2 - a}{3}$, where $ma = [RTS]$.

Showing the first property is quite simple

$$1 + \alpha_1 + \alpha_2 + \alpha_3 = \sum_{x \in \mathbb{F}_q} \omega^x = \left(\sum_{x_1 \in \mathbb{F}_p} \omega_1^{x_1} \right) \cdots \left(\sum_{x_k \in \mathbb{F}_p} \omega_k^{x_k} \right) = 0.$$

To show the second property, let us introduce integers b, c which satisfy $mb = [SST], mc = [STT]$. Now, we compute the product $\alpha_1\alpha_2$ as

$$\alpha_1\alpha_2 = \sum_{x \in R} \sum_{y \in S} \omega^{x+y} = \sum_{x \in {\mathbb{F}_q}^\times} [RS\{-x\}]\omega^x = \sum_{x \in R} [RS\{-x\}]\omega^x + \sum_{x \in S} [RS\{-x\}]\omega^x + \sum_{x \in T} [RS\{-x\}]\omega^x.$$

Note that $[RS\{-x\}]$ is constant over all $x \in R$, say, since the scaling property $[RS\{-x\}] = [rR, rS, \{-rx\}] = [RS\{-rx\}]$ for any $r \in R$. Hence, we have $[RS\{-x\}] = \frac{1}{m}[RSR] = \frac{1}{m}[SST] = b$ for all $x \in R$, and similarly $[RS\{-x\}] = c$ for $x \in S$, $[RS\{-x\}] = a$ for $x \in T$. Hence,

$$\alpha_1\alpha_2 = b \sum_{x \in R} \omega^x + c \sum_{x \in S} \omega^x + a \sum_{x \in T} \omega^x = b\alpha_1 + c\alpha_2 + a\alpha_3.$$

In a similar way, one can derive $\alpha_2\alpha_3 = a\alpha_1 + b\alpha_2 + c\alpha_3$ and $\alpha_3\alpha_1 = c\alpha_1 + a\alpha_2 + b\alpha_3$. Combined with the first property, summing these three relations gives the second property directly. Finally, we have

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

Noting that $a + b + c = \frac{1}{m}\left([ST\mathbb{F}_q] - [ST\{0\}]\right) = m$ completes the proof of the third property. Let us now define the integers $A = 9a - 3m - 2$ and $B = b - c$. The goal will be to show $A^2 + 27B^2 = 4q$, which will complete the proof immediately, since

$$\#C/\mathbb{F}_q = \frac{9[RTS]}{m} = 9a = A + 3m + 2 = q + 1 + A.$$

To show $4q = A^2 + 27B^2$, we consider polynomials $f(X) = (X - \alpha_1)(X - \alpha_2)(X - \alpha_3)$ and $g(X) = (X - \beta_1)(X - \beta_2)(X - \beta_3)$, where $\beta_i = 1 + 3\alpha_i$. It is not hard to verify the properties $\beta_1 + \beta_2 + \beta_3 = 0, \beta_1\beta_2 + \beta_2\beta_3 + \beta_3\beta_1 = -3p, \beta_1\beta_2\beta_3 = Aq$. Vieta formulas now give

$$f(X) = X^3 + X^2 - mX - am + \frac{m^2 - a}{3}, \qquad g(X) = X^3 - 3qX + Ap.$$

Note that the discriminants of f, g, denoted by Δ_f, Δ_g, can be satisfy

$$\Delta_g = \prod_{i \neq j} (\beta_i - \beta_j) = 3^6 \prod_{i \neq j} (\alpha_i - \alpha_j) = 3^6 \Delta_f.$$

The discriminant of f can be computed directly through the following calculation

$$\begin{aligned}\Delta_f^{1/2} &= \prod_{i<j}(\alpha_i - \alpha_j) = \alpha_1\alpha_2(\alpha_1 - \alpha_2) + \alpha_2\alpha_3(\alpha_2 - \alpha_3) + \alpha_3\alpha_1(\alpha_3 - \alpha_1) \\ &= (b-c)(\alpha_1^2 + \alpha_2^2 + \alpha_3^2 - \alpha_1\alpha_2 - \alpha_2\alpha_3 - \alpha_3\alpha_1) = Bq.\end{aligned}$$

On the other hand, Δ_g can be computed from the coefficients of g, which yields

$$\Delta_g = 27(4q^3 - A^2q^2).$$

Recalling that $\Delta_g = 27^2\Delta_f$, we obtain $4q^3 - A^2q^2 = 27B^2q^2$, giving us exactly $q = A^2 + 27B^2$. This completes the proof. □

1.6 History

In this section, we present a historical account of the theory of zeta functions of algebraic curves, as presented by Roquette in [23] and Milne in [20]. Roquette's book covers the development of the theory of zeta functions over function fields, from Artin's work in the 1920s to Weil's proofs of Artin's conjectures in 1940. On the other hand, Milne focuses on the story of Weil conjectures, which were a generalization of Artin's conjectures to arbitrary varieties, and focuses on the development which led to Deligne's proof of the Riemann Hypothesis in 1974.

The history of the investigation of the zeta function of function fields starts with Artin. In his thesis [2], he investigated quadratic function fields of positive characteristic and defined a notion of zeta function for them. He shows that the zeta function is rational and satisfies the functional equation, which he then uses to derive the analogue of the trace formula (1.3) for the function fields. Then, Artin goes on to conjecture the statement analogous the Riemann Hypothesis for his zeta function of function fields in one variable.

Artin's results on rationality and functional equation of the zeta function were then generalized by F.K. Schmidt in 1931 [24], who built the theory of function fields needed to show the Riemann-Roch theorem and used it to show Artin's conjectures, in a very similar way as presented in Chapter 4.

After Artin verified the Riemann Hypothesis for quadratic function fields, in 1933 Hasse [15] gave a proof of the Riemann Hypothesis for elliptic function fields (which correspond to elliptic curves), using the theory of complex multiplication on elliptic curves.[3] This was one of the first instances where the basics of geometric intuition started being used instead of algebraic ones. Even today the bound on the number of points on elliptic curves is known as the Hasse bound.

In 1940 and 1941, Weil [34], [35] approached this problem from the perspective of algebraic curves and announced the proofs of the Riemann Hypothesis for curves. Weil's work involved a shift in perspective from considering the zeta functions purely algebraically towards introducing geometric reasoning. Due to the turbulent times he was living in, Weil hurried to make the announcement of his proof, and he had to work for several more years before publishing the final version of his argument. Formalizing these proofs was one of the main driving factors behind Weil's work on foundations of algebraic geometry in the 1940s.

Weil came up with two essentially different approaches to showing the Riemann Hypothesis for a curve $C/\mathbb{F}_q$. The first one considered the surface $C \times C$ (somewhat similar to Bombieri's argument presented in Chapter 6), and considered a intersection of the diagonal $\Delta = \{(P, P) : P \in C\}$ with the curve $C_1 = \{(P, P^q) : P \in C\}$, where $P \mapsto P^q$ denoted the Frobenius morphism. Then, using the intersection theory on surfaces he developed, Weil managed to bound the number of intersections between these two curves and obtain the proof of the Riemann Hypothesis. This argument was presented in [36].

The second Weil's proof [37] relied on a geometric construction known as the Jacobian of the curve. Although the proof begins in the same way, the number of intersections between Δ and C_1 is now bounded by passing to the Jacobian of the curve, somewhat similarly to Hasse's proof in the case of elliptic curves.

While proving Artin's conjectures for curves, Weil formulated a set of conjectures describing the behavior of zeta functions of varieties, connecting the rational point counts to topological properties of the curve, such

[3] According to Roquette, Hasse never published the full proof of his bound, but only the outline in the paper we cite.

as the genus and the Betti numbers. With the state of algebraic geometry in the 1940s, these conjectures were well beyond reach. However, these conjectures served as a productive inspiration for the further development of modern algebraic geometry.

In parallel to the resolution of the Weil conjectures for general varieties, the Riemann Hypothesis for curves also saw renewed interest during the 1960s and 1970s. Namely, Stepanov's work in the 1960s [28] showed that the polynomial method may be used to derive the Riemann Hypothesis for some special cases of curves, such as hyperelliptic curves. This came as a surprise, but the method was soon extended by W. M. Schmidt and Bombieri to show the Riemann Hypothesis for all curves.

Since its introduction, the Stepanov method has found other use cases, unrelated to the zeta functions of algebraic curves. For example, variants of the Stepanov method have been used by Bourgain, Gamburd, and Sarnak [5] in the investigation of the Markoff equation, where directly applying the bounds coming from Weil's theorem was not sufficient, and by Heath-Brown, Konyagin [16] and Bourgrain, Konyagin [6] to obtain improved bounds on various exponential sums.

Chapter 2

Elements of Algebraic Geometry

In this section, we will introduce the basic tools of algebraic geometry of curves, with the goal of proving several basic results which we will use later in the proof of the Riemann-Roch theorem.

Let us begin by discussing the notions of affine and projective spaces, which will be ubiquitous throughout the rest of the presentation. The affine plane over $\mathbb{F}_q$ is nothing else than ${\mathbb{F}_q}^2$.

The projective plane over $\mathbb{F}_q$, denoted by $\mathbb{P}^2(\mathbb{F}_q)$, will be constructed as a quotient of ${\mathbb{F}_q}^3 - (0,0,0)$ and a certain equivalence relation. Namely, we define the equivalence relation by declaring the points $(x,y,z),(x',y',z') \in {\mathbb{F}_q}^3$ to be equivalent if there exists a nonzero scalar $\lambda \in \mathbb{F}_q$ such that $x = \lambda x', y = \lambda y'$ and $z = \lambda z'$. Then, one can define $\mathbb{P}^2(\mathbb{F}_q)$ to be the set of equivalence classes of points $(x,y,z) \in {\mathbb{F}_q}^3$ under this equivalence relation. We will denote the equivalence class of (x,y,z) by $[x:y:z]$.

In a similar way as above, one can define the projective space $\mathbb{P}^2(\overline{\mathbb{F}_q})$, which contains $\mathbb{P}^2(\mathbb{F}_{q^m})$ as a subset for all m. Let us note a slight subtlety regarding the interaction between $\mathbb{P}^2(\overline{\mathbb{F}_q})$ and $\mathbb{P}^2(\mathbb{F}_{q^m})$. Suppose we have a point $[x:y:z] \in \mathbb{P}^2(\overline{\mathbb{F}_q})$ and ask whether this point lies in $\mathbb{P}^2(\mathbb{F}_{q^m})$. A natural answer would be to say that $[x:y:z] \in \mathbb{P}^2(\mathbb{F}_{q^m})$ if and only if the coordinates of the point are elements of $\mathbb{F}_{q^m}$, i.e. $x,y,z \in \mathbb{F}_{q^m}$. But since points of $\mathbb{P}^2(\overline{\mathbb{F}_q})$ may be scaled using arbitrary scalars $\lambda \overline{\mathbb{F}_q}$, we may scale the coordinates of a point of $\mathbb{P}^2(\mathbb{F}_{q^m})$ so that they do not lie in $\mathbb{F}_{q^m}$. Hence, we need to be more careful, and therefore, we say that a point $[x:y:z] \in \mathbb{P}^2(\overline{\mathbb{F}_q})$ is defined over $\mathbb{F}_{q^m}$ if there is a nonzero scalar $\lambda \in \overline{\mathbb{F}_q}$ for which $\lambda x, \lambda y, \lambda z \in \mathbb{F}_{q^m}$.

As we will see in the next section too, the projective and affine planes are closely related. Namely, the affine plane ${\mathbb{F}_q}^2$ can be identified with a subset of the projective plane $\mathbb{P}^2(\mathbb{F}_q)$ in a natural way, by sending $(x,y) \in {\mathbb{F}_q}^2$ to $[x:y:1] \in \mathbb{P}^2(\mathbb{F}_q)$. Moreover, it is interesting to note that one can cover the projective plane using three affine planes in this way, although we will not need this fact in our presentation.

2.1 Rational functions on curves

In this section, we will define the rational functions on a plane curve $C/\mathbb{F}_q$, which may be either affine or projective curves. We will start by discussing the affine case. As before, we will always assume that $C/\mathbb{F}_q$ is a smooth plane curve defined by an absolutely irreducible polynomial $f(X,Y)$ in the affine case or $F(X,Y,Z)$ in the projective case.

Definition 2.1. If $C/\mathbb{F}_q$ is an affine curve, a *rational function* on $C/\mathbb{F}_q$ is a ratio of polynomials $\frac{g}{h}$ where $g,h \in \mathbb{F}_q[X,Y]$ are polynomials and h is not divisible by f.

In fact, the set of rational functions on a curve $C/\mathbb{F}_q$ can be thought of as a field in the following way. If we define the addition and multiplication of rational functions in the usual way, and we say that two rational functions $\frac{g_1}{h_1}, \frac{g_2}{h_2}$ are equal if $g_1 h_2 - h_1 g_2$ is divisible by f, then the set of rational functions becomes a field. This field, denoted by $\mathbb{F}_q(C)$, is precisely the field of fractions of the quotient ring $\mathbb{F}_q[X,Y]/(f)$ (in fact, we can take this as a definition of $\mathbb{F}_q(C)$). We will call $\mathbb{F}_q(C)$ the *function field* of the curve $C/\mathbb{F}_q$, and it will be one of the main objects we will use to study the curve $C/\mathbb{F}_q$. [1] Note that every rational function $\varphi = \frac{g}{h} \in \mathbb{F}_q(C)$

[1] There exists yet another way to define the function field, which involved appealing to more complicated algebraic construction

can be evaluated at all point $P = (x, y) \in C/\overline{\mathbb{F}_q}$ which have $h(x, y) \neq 0$, simply by setting $\varphi(P) = \frac{g(x,y)}{h(x,y)}$. One might worry whether such evaluation map is well defined, given that we are considering g and h only modulo f, but since f vanishes at all points $(x, y) \in C/\overline{\mathbb{F}_q}$ this is not an issue. Finally, note that the above definitions and discussion extend seamlessly to define the rational functions on $C/\overline{\mathbb{F}_q}$.

The definitions may be extended to the case of projective curves in almost the same way. However, one should be careful and require that the numerator and denominator of the rational function have the same degree, in order to be able to evaluate this rational function on the points of the curve.

Definition 2.2. If $C/\mathbb{F}_q$ is a projective plane curve, a *rational function* on $C/\mathbb{F}_q$ is a ratio of polynomials $\frac{G}{H}$ where $G, H \in \mathbb{F}_q[X, Y, Z]$ are homogeneous polynomials of the same degree and F does not divide H.

Let us give a brief comment about the above definition and in particular about the additional requirement that G and H should be homogeneous polynomial of the same degree, say $\deg G = \deg H = d$. If we recall that we are working in projective space, where $P = [x : y : z] = [\lambda x : \lambda y : \lambda z]$, we see that this requirement ensures that the rational function G/H evaluates to the same value on the points $[x : y : z]$ and $[\lambda x : \lambda y : \lambda z]$, simply because

$$\frac{G(\lambda x, \lambda y, \lambda z)}{H(\lambda x, \lambda y, \lambda z)} = \frac{\lambda^d G(x, y, z)}{\lambda^d H(x, y, z)} = \frac{G(x, y, z)}{H(x, y, z)}.$$

Hence, if $\varphi = \frac{G}{H}$, we may define $\varphi(P) = \frac{G(x,y,z)}{H(x,y,z)}$. By the above discussion, this definition gives a well-defined map from $C/\overline{\mathbb{F}_q} - \{H(x, y, z) = 0\}$ to $\overline{\mathbb{F}_q}$.

Of course, one can make the set of rational function on C a field, as previously discussed in the affine case, by adding and multiplying the rational functions in the usual way, along with considering them modulo F.

Given a point P, we will now focus on the set of functions which can be evaluated at P.

Definition 2.3. If $C/\mathbb{F}_q$ is an affine plane curve and $P \in C/\overline{\mathbb{F}_q}$, we call a rational function $\varphi \in \overline{\mathbb{F}_q}(C)$ is *regular at a point* $P \in C/\overline{\mathbb{F}_q}$ if it can be written as a ratio $\varphi = \frac{g}{h}$, where $g, h \in \mathbb{F}_q[X, Y]$ and $h(P) \neq 0$.

The set of regular functions at a given point P forms a ring, which we call the *local ring* and denote by $\mathcal{O}_P$. Note that the local ring at P contains the rational functions $\varphi \in \overline{\mathbb{F}_q}(C)$ and not only in $\mathbb{F}_q(C)$. We define the *maximal ideal of* C at P is $\mathfrak{M}_P = \{\varphi \in \mathcal{O}_P : \varphi(P) = 0\}$. It is not hard to check that $\mathfrak{M}_P$ is indeed a maximal ideal in $\mathcal{O}_P$. Note that the above definitions extend seamlessly to the case of projective plane curves.

We will finish this section with an explanation how to relate rational functions of an affine and projective curve. Namely, if an affine curve $C/\mathbb{F}_q$ was defined by the equation $f(X, Y) = 0$, one can produce the projective version of this curve by homogenizing the polynomial $f(X, Y)$. More precisely, if $\deg f = d$ and $f(X, Y) = \sum_{i+j \leq d} a_{ij} X^i Y^j$, we define the *homogenization* of f to be the homogeneous polynomial $F(X, Y, Z) = \sum_{i+j \leq d} a_{ij} X^i Y^j Z^{d-i-j}$. Then, F defines a projective curve $\widetilde{C}/\mathbb{F}_q \subset \mathbb{P}^2(\mathbb{F}_q)$.

The described procedure shows how to start from an affine curve $C/\mathbb{F}_q$ and produce a projective curve $\widetilde{C}/\mathbb{F}_q$, but one can also do the reverse. Note that when we restrict the curve $\widetilde{C}/\mathbb{F}_q$ to the affine plane of $\mathbb{P}^2(\mathbb{F}_q)$ given by $\{[x : y : z] \in \mathbb{P}^2(\mathbb{F}_q) : z \neq 0\}$, we obtain the affine curve $C/\mathbb{F}_q$, which is precisely the curve we started with. Moreover, the defining equation of the affine curve is given by $f(X, Y) = F(X, Y, 1) = 0$.

In order to be able to refer to this correspondence later on, we state it formally in the following proposition, which is easily proved directly from the definitions.

Proposition 2.4. Let $\widetilde{C}/\overline{\mathbb{F}_q} \subset \mathbb{P}^2(\overline{\mathbb{F}_q})$ be a smooth projective curve defined by $F(x, y, z) = 0$, and let $A_0 = \{[x : y : z] \in \mathbb{P}^2(\overline{\mathbb{F}_q}) | z \neq 0\}$ be an affine plane within $\overline{\mathbb{P}}^2(\mathbb{F}_q)$. Then,

- If $C = \widetilde{C} \cap A_0$, then C is a smooth affine plane curve given by $f(x, y) = F(x, y, 1) = 0$.
- If P is a point of $C/\overline{\mathbb{F}_q}$ (and thus also of $\widetilde{C}/\overline{\mathbb{F}_q}$), the local rings of $\widetilde{C}$ and C at P are isomorphic, and the isomorphism is given by $\frac{G(x,y,z)}{H(x,y,z)} \mapsto \frac{G(x,y,1)}{H(x,y,1)}$. Moreover, the same isomorphism shows that the function fields $\mathbb{F}_q(C)$ and $\mathbb{F}_q(\widetilde{C})$ are isomorphic.

called *localization*. Although we will not define it now, let us simply mention that $\mathbb{F}_q(C)$ is isomorphic to the quotient of the local ring $\mathbb{F}_q[X, Y]_{(f)}$ with its maximal ideal.

2.2 Valuations

In this section, we introduce one of the basic algebraic constructions which will be used throughout the thesis, namely valuations. They will play a key role in defining the order of vanishing of a function at a point. We also introduce discrete valuation rings, with the goal of showing that the ring of regular functions $\mathcal{O}_P$ is a DVR. This will be the topic of the next section.

Definition 2.5. A ring R is a *discrete valuation ring* (DVR) is there exists an element $t \in R$ such that every element $r \in R$ can be uniquely written as $r = ut^n$ for a unit u. The element t is called the *uniformizer*.

Example 2.6. The following example has the goal of elucidating the intuition behind the definition of DVRs, as well as introducing some ideas which will be used in the subsequent proofs. Consider the set of rational functions $S \subset \mathbb{C}(x)$ which do not have a pole at $x = 1$. We can easily associate the order of vanishing at 1 to every function $\varphi \in S$ by simply looking at the highest power of $(x-1)$ dividing φ. In other words, we may represent every $\varphi \in S$ as $\varphi = (x-1)^n\psi$, where $\psi(1) \neq 0$. Since $\psi(1) \neq 0$, we have $\psi^{-1} \in S$ too and ψ is a unit. Hence, we conclude that φ is indeed expressible as a product of $(x-1)^n$ and a unit, which corresponds exactly to the definition of DVRs. From this perspective, it seems natural to define the analogue of the order of vanishing, the so called *valuation*, of an element $r \in R$ to be the power of t appearing in its expression as $r = ut^n$.

As we saw in the previous example, DVRs are often constructed by starting from a field, and selecting a set of elements satisfying a certain property (such as not having a pole at $x = 1$). Let us now formalize this process and introduce what a valuation is.

Definition 2.7. If K is a field, a *valuation* on K is a surjective function $\nu : K \to \mathbb{Z} \cup \{\infty\}$ satisfying $\nu(xy) = \nu(x) + \nu(y)$, $\nu(x+y) \geq \min\{v(x), v(y)\}$ for all $x, y \in K$ and $\nu(x) = \infty$ if and only if $x = 0$.

Let us now explain how, starting from a DVR R, one can construct a field with a valuation. If a ring R is a DVR with a field of fractions K, then K has a natural valuation on it. Recall that every element of $r \in K$ can be written as $r = t^n u$ for some $n \in \mathbb{Z}$ and a unit u. Then, we may define $\nu(r) = n$. Again, it is not hard to verify that ν satisfies the properties of a valuation.

This procedure can also be reversed. If K is a field with a valuation ν, we can define the *valuation ring* R as the set of element of nonnegative valuation, i.e. $R = \{r \in K : \nu(r) \geq 0\}$. It is not hard to see that R is a DVR, and that its uniformizer can be chosen to be any element with $\nu(t) = 1$ (which exists by surjectivity of ν). The set of elements of degree ≥ 1 will then be the maximal ideal of the valuation ring. As elementary abstract algebra shows, the quotient of a ring by its maximal ideal gives a field, and hence we define the *residue field* of a valuation ν is simply the quotient of the valuation ring by its maximal ideal.

Example 2.8. The simplest example of a valuation which is very useful is the p-adic valuation. Namely, if p is a prime number, we can define the valuation ν_p on the rational numbers by setting

$$\nu_p\left(\frac{a}{b}\right) = \text{highest power of } p \text{ dividing } a - \text{highest power of } p \text{ dividing } b.$$

Checking that ν_p is indeed a valuation follows from elementary number theoretic properties of divisibility. The residue field of this valuation is $\mathbb{F}_p$. We will encounter this valuation again in Chapter 5, where we will use it extensively in order to define the notion of hyperderivatives.

In order to get more familiar with the properties of valuations, let us record the following useful property of valuations, called the *strict triangle inequality*. Although we will not use this property in this section, it will come in very handy later on.

Proposition 2.9. Let ν be a valuation on the field K. If $x, y \in K$ are elements with $\nu(x) < \nu(y)$, then $\nu(x+y) = \nu(x)$.

Proof. By the triangle inequality, we know that $\nu(x+y) \geq \min\{\nu(x), \nu(y)\} = \nu(x)$. If we had $\nu(x+y) > \nu(x)$, we would also have $\nu(x) \geq \min\{\nu(x+y), \nu(-y)\} > \nu(x)$, which is not possible. Hence, $\nu(x+y) = \nu(x)$. □

Now that we have introduced the general notion of DVRs and valuations, let us focus on showing that $\mathcal{O}_P$ is a DVR. Note that the field of fractions of $\mathcal{O}_P$ is indeed the set of all rational functions $\overline{\mathbb{F}_q}(C)$. Hence, showing that $\mathcal{O}_P$ is a DVR implies, by the above discussion, that we can define a valuation associated to P, denoted by ν_P, which will represent the order of zero/pole of a function at P.

Proposition 2.10. Let $C/\overline{\mathbb{F}_q}$ be a smooth curve and let P be a point on $C/\overline{\mathbb{F}_q}$. The ring of regular functions at P, $\mathcal{O}_P$, is a DVR.

We will show this result in the next section. However, before that we need to develop the algebraic machinery needed for it. Hence, we divert our attention to Noetherian rings, local rings and other basic objects of commutative algebra. The main reason for this is that we will show that $\mathcal{O}_P$ is a DVR by showing it is Noetherian, local and that its maximal ideal is principal.

Definition 2.11. A ring R is *Noetherian* if every ascending chain of ideals has a maximal element. In other words, if $I_1 \subseteq I_2 \subseteq \dots$ is a chain of ideals, then we have $I_n = I_{n+1}$ for all big enough n. A ring R is *local* if it has only one maximal ideal. A ring R is a *domain* if it has no zero-divisors, i.e. if $rs = 0$ implies either $r = 0$ or $s = 0$.

Proposition 2.12. The ring of regular functions at P, $\mathcal{O}_P$, is a Noetherian ring.

Let us outline the strategy to prove Proposition 2.12. Since we are working with local rings, we may assume our curve is affine. We will begin by showing that $\mathcal{O}_P$ is a localization of the ring $\mathbb{F}_q[X,Y]/(f)$, and then we will show that localization of Noetherian rings are Noetherian. The only thing which will remain then is to show that $\mathbb{F}_q[X,Y]/(f)$ is Noetherian, which will be shown through the combination of two facts - that $\mathbb{F}_q[X,Y]$ is Noetherian and that quotients of Noetherian rings are Noetherian.

In order to carry though with this proof, we need to introduce the notion of localization, since thinking of $\mathcal{O}_P$ as a localization will be immensely helpful for proving it is Noetherian. Our discussion of localization is based on [10].

Let R be a domain, and let $\mathfrak{p} \subset R$ be a prime ideal. Then, we define the localization of R at $\mathfrak{p}$, denoted by $R_{(\mathfrak{p})}$ as the set of fractions $\{\frac{a}{b}, b \notin \mathfrak{p}\}$, where we declare two fractions $\frac{a_1}{b_1}$ and $\frac{a_2}{b_2}$ equal if $b_2a_1 - a_2b_1 \in \mathfrak{p}$. It is not hard to check that $R_{\mathfrak{p}}$ is indeed a ring under the usual addition and multiplication operations. Note that we can think of R as a subset of $R_{(\mathfrak{p})}$, since every element $a \in R$ can be associated with $a/1 \in R_{(\mathfrak{p})}$.

From this perspective, it is not hard to see that $\mathcal{O}_P$ is a localization of the ring $\mathbb{F}_q[X,Y]/(f)$ at the maximal ideal $\mathfrak{M}_P$. Hence, in order to talk about the localization being Noetherian, we need to understand the ideals of the localization.

Lemma 2.13. Let R be a domain, let $R_{(\mathfrak{p})}$ be its localization at the prime ideal $\mathfrak{p}$ and let $J \subseteq R_{(\mathfrak{p})}$ be an proper ideal. Then there exists an ideal I of R with $I \subset \mathfrak{p}$ such that $J = \{\frac{i}{s} : i \in I, s \notin \mathfrak{p}\}$.

Proof. We construct I by setting $I = J \cap R$. We claim that $I \subset \mathfrak{p}$. Suppose we had an element $i \in I$ with $i \notin \mathfrak{p}$. Then $i \in J$ and $1/i \in R_{(\mathfrak{p})}$ implies that $i \cdot 1/i = 1 \in J$, which is a contradiction to the assumption J is not proper. Hence $I \subset \mathfrak{p}$.

Let us now show $J = \{\frac{i}{s} : i \in I, s \notin \mathfrak{p}\}$. If we pick an element $j \in J$, we have $j = i/s$, for some $i \in A$ and $s \notin \mathfrak{p}$. But we also have $s \in R_{(\mathfrak{p})}$ and so $js = i \in J$. Since $i \in A$ and $i \in J$ we conclude $i \in I$ and $j \in \{\frac{i}{s} : i \in I, s \notin \mathfrak{p}\}$.

To show the other inclusion, let us pick $i \in I, s \notin \mathfrak{p}$. Then $i \in J$ and $1/s \in R_{(\mathfrak{p})}$, meaning that $i/s = i \cdot 1/s \in J$. This completes the proof. □

Now, it is almost obvious that localizations of Noetherian rings are Noetherian.

Lemma 2.14. If R is a Noetherian ring and $\mathfrak{p}$ its prime ideal, the ring $R_{(\mathfrak{p})}$ is also Noetherian.

Proof. Let $J_1 \subseteq J_2 \subseteq \dots$ be an ascending chain of ideals of $R_{(\mathfrak{p})}$. By Lemma 2.13 we know that there are corresponding ideals $I_1 \subseteq I_2 \subseteq \dots$ satisfying $J_k = \{\frac{i}{s} : i \in I_k, s \notin \mathfrak{p}\}$. Since R is Noetherian, we know that for all big enough n we have $I_n = I_{n+1}$. But then we also have $J_n = J_{n+1}$, showing that every ascending chain of ideals stabilizes. □

As described in the outline of our proof, we now show that $\mathbb{F}_q[X,Y]$ is Noetherian, and that quotients of Noetherian rings are Noetherian.

Lemma 2.15. Let R be a domain and let $I \subset R$ be an ideal. Then R/I is a Noetherian ring.

Proof. This proof is almost the same as the proof of Lemma 2.14. We will use the elementary fact from abstract algebra that the ideals of R/I correspond to ideals of R containing I, where the correspondence is given by $J \mapsto J + I$.If $J_1 \subseteq J_2 \subseteq \dots$ is an ascending chain of ideals of R/I, we use the above fact to find a chain of ideals $I_1 \subseteq I_2 \subseteq \dots$ satisfying $I_k/I = J_k$. Since R is Noetherian, we know that for all big enough n we have $I_n = I_{n+1}$. But then we also have $J_n = J_{n+1}$, showing that every ascending chain of ideals stabilizes. □

Finally, the last step will be to show that $\mathbb{F}_q[X,Y]$ is Noetherian. For this, we use the Hilbert basis theorem.

Lemma 2.16. If R is a Noetherian ring, so is the polynomial ring $R[X]$.

Proof. This proof follows the approach presented in [17]. Let $J_1 \subseteq J_2 \subseteq \cdots$ be an ascending chain of ideals in $R[X]$ and let

$$I_{k,\ell} = \{a \in A | \text{there is } f \in J_k \text{ with degree } \ell \text{ and leading coefficient } a\},$$

for $k \geq 1, \ell \geq 0$ Since J_k are ideals of $R[X]$ for $k \geq 1$, we easily see that $I_{k\ell}$ are also ideals of R. Moreover, the ideals $I_{k,\ell}$ form an ascending chain in both coordinates. This is not hard to see: for $k \leq k'$ we have $I_{k,\ell} \subseteq I_{k',\ell}$ since $J_k \subseteq J_{k'}$ and for $\ell \leq \ell'$ we have $I_{k,\ell} \subseteq I_{k1,\ell'}$ since J_k is closed under multiplication by X.

Since R is Noetherian, the ascending chain of ideals $I_{k,0} \subseteq I_{k,1} \subseteq \cdots$ stabilizes and we can write $I_{k,\ell} = \mathfrak{m}_k$ for all big enough ℓ. Applying the Noetherian property again, we find that the chain $\mathfrak{m}_1 \subseteq \mathfrak{m}_2 \subseteq \cdots$ stabilizes and that for some A and all $k \geq A$ we have $\mathfrak{m}_k = \mathfrak{m}_A$. Note that we also have $I_{A,0} \subseteq I_{A,1} \subseteq \cdots$ stabilizes at $\mathfrak{m}_A$, and so we may write $\mathfrak{m}_A = I_{A,B}$.

The goal is now to show that we may increase A such that $I_{k,\ell} = I_{A,\ell}$ for all $k \geq A$ and all $\ell \geq 0$. Note that for $\ell \geq B$, we have $I_{k,\ell} \subseteq I_{A,B} \subseteq I_{A,\ell} \subseteq I_{k,\ell}$, so we indeed have $I_{k,\ell} = I_{A,\ell}$. On the other hand, for each $\ell < B$, the ascending chain $I_{0,\ell} \subseteq I_{1,\ell} \subseteq \cdots$ stabilizes at a finite index A_ℓ. Increasing A to obtain $A > \max_{\ell \in [0,B]} A_\ell$ now suffices to ensure that $I_{k,\ell} = I_{A,\ell}$ for all $k \geq A$ and all $\ell \geq 0$.

Now, we claim that $J_1 \subseteq J_2 \subseteq \cdots$ stabilizes at J_A, i.e. that we have $J_k = J_A$ for all $k \geq A$. The inclusion $J_A \subseteq J_k$ is clear, and our goal is to prove the reverse inclusion, i.e. to show that for every polynomial $f \in J_k$ we also have $f \in J_A$. We do this by the induction on the degree of the polynomial $f \in J_k$.

If f has degree 0, we have $f \in R$ and so $f \in I_{k,0} \subseteq I_{A,0} \subseteq J_A$. To perform the induction step, let f be a polynomial of degree ℓ with leading coefficient a and note $a \in I_{k,\ell}$. We then have $a \in I_{A,\ell}$ and so there exists a polynomial $g \in J_A$ with degree ℓ and leading coefficient a. Since $J_A \subseteq J_k$ we also have $g \in J_k$ and so $f - g \in J_k$. Note that $f - g$ has degree $\ell - 1$ and by the induction hypothesis we obtain $f - g \in J_A$. Since $g \in J_A$, this also implies $f \in J_A$, and we conclude that $J_k \subseteq J_A$. This completes the proof. □

Proof of Proposition 2.12. The proof follows directly from the algebraic machinery that we have introduced. Since we are working with local ring, we may assume we are working with an affine rather than a projective curve.

We have $\mathbb{F}_q$ is a field and hence a Noetherian ring. By applying Hilbert's basis theorem (i.e. Lemma 2.16) twice, we obtain that $\mathbb{F}_q[X,Y]$ is Noetherian. Then, by Lemma 2.15, we conclude that $\mathbb{F}_q[X,Y]/(f)$ is a Noetherian ring. Further, the local ring $\mathcal{O}_P$ is a localization of $\mathbb{F}_q[X,Y]/(f)$ at the maximal ideal $\mathfrak{M}_P$, and by Lemma 2.14 it is also Noetherian. This completes the proof. □

Proposition 2.17. If a ring R is Noetherian, local domain whose maximal ideal is principal, then R is a DVR.

Proof. Let t be the generator of the maximal ideal $M \subset R$. We start from the observation that M contains all non-unit elements of R. To see this, let r be a non-unit element, and consider the ideal (r) generated by it. This ideal is contained in a maximal one, by the fact R is Noetherian, and hence it must be contained in M. In particular, r is contained in M.

Now, let us pick $r \in R$ and show how to produce a representation $r = ut^n$. If r is a unit, the claim is obvious. Otherwise, we have $r \in M$ and so $r = tr_1$. If r_1 is a unit, the claim is proven and otherwise we have

$r_1 = tr_2$. This process continues to produce a sequence $(r) \subseteq (r_1) \subseteq (r_2) \subseteq \dots$. If any of the elements r_n is a unit, we have the representation $r = t^n r_n$. Otherwise, we obtain an infinite ascending chain of ideals, which must have a maximal element since R is Noetherian. In other words we have $(r_n) = (r_{n+1})$ for some n. But this means that the r_n and r_{n+1} differ by a unit u, i.e. we have $r_n = ur_{n+1}$. However, recall that we have $r_n = tr_{n+1}$, and so $t = u$ must be a unit. But this is a contradiction since M is a proper ideal. Hence, the described process ends and we have $r = ut^n$ for some $u \in R, n \in \mathbb{Z}_{\geq 0}$. Once the existence of the representation has been established, the uniqueness is clear. □

2.3 Rings of regular functions as DVRs

In this section, we show the proof of Proposition 2.10, using the algebraic machinery developed in the Section 2.2.

Equipped with this result, we are finally ready to prove the main result of this section, Proposition 2.10.

Proof of Proposition 2.10. Our exposition follows Fulton's approach from [11]. Using the strategy established in Section 2.2, we only need to check that $\mathcal{O}_P$ satisfies the conditions of Proposition 2.17. That $\mathcal{O}_P$ is Noetherian we have already checked in Proposition 2.12. Note that P belongs to some affine chart of $\mathbb{P}^2(\mathbb{F}_q)$, and hence we may restrict the curve C to this affine subset giving us an affine curve.

Furthermore, it is not hard to check that $\mathcal{O}_P$ is a local domain. The fact $\mathcal{O}_P$ is a domain follows from the irreducibility of the defining equation for the curve C. To show $\mathcal{O}_P$ is local, we note that the maximal ideal $\mathfrak{M}_P \subset O_P$ consists of all functions vanishing at P, and if a function $\frac{g}{h}$ does not vanish at P, i.e. we have $g(P) \neq 0$, then we have $\frac{h}{g} \in \mathcal{O}_P$ and so $\frac{g}{h}$ is a unit. Hence, all non-units of $\mathcal{O}_P$ belong to the maximal ideal $\mathfrak{M}_P$.

Finally, we need to check that the $\mathfrak{M}_P$ is a principal ideal. This is the key part of the proof, where we make use of the smoothness of the curve $C/\overline{\mathbb{F}_q}$. To simplify the rest of the proof, we will assume our curve is affine and that P lies at the origin. Furthermore, since C is smooth at P, we will assume that $\frac{\partial f}{\partial X}(0,0) = 0$ and $\frac{\partial f}{\partial Y} = 1$. This is not a loss of generality since this may be ensured through an affine change of coordinates, which does not affect the structure of $\mathcal{O}_P$.

Now, the ideal $\mathfrak{M}_P = \{\frac{g}{h} : g(0,0) = 0, h(0,0) \neq 0, g, h \in \overline{\mathbb{F}_q}[X,Y]/(f)\}$ can be generated by elements X and Y. The reason for this is that every polynomial $g \in \mathbb{F}_q[X,Y]$ satisfying $g(0,0) = 0$ can be written as $g = Xg_1 + Yg_2$, and consequently every element $\frac{g}{h}$ of $\mathfrak{M}_P$ can be written as $X\frac{g_1}{h} + Y\frac{g_2}{h}$. To prove that $\mathfrak{M}_P$ is principal, we will show that Y can be expressed as $\frac{g'}{h'}X^2$, for some $\frac{g'}{h'} \in \mathcal{O}_P$.

The key trick we will use to express Y in this form is that we are working modulo f. More precisely, since $\partial_X f(0,0) = 0$ and $\partial_Y f(0,0) = 1$, we have $f(X,Y) = Y + (\text{terms of degree} \geq 2)$. The higher degree terms can be split into those containing X^2 and the rest, giving us the following expression $f = Y(1+h) + X^2 g$, where $h(0,0) = 0$. Hence, we have $Y = \frac{g}{1+h}X^2 \pmod f$ and hence $\mathfrak{M}_P$ is generated solely by X. This completes the proof and shows that $\mathcal{O}_P$ is a DVR. □

Remark 2.18. The last part of this proof actually shows that not only is $\mathfrak{M}_P$ a principal ideal, but it is generated by the equation of any line through $P = (x,y) \in C$ which is not tangent to the curve C. More precisely, if $g(X,Y) = a(X-x) + b(Y-y)$ defines a line through P, we have that g is a uniformizer for the ring $\mathcal{O}_P$ if $[a:b] \neq [\partial_X f : \partial_Y f]$. By applying an affine transformation, one can easily convert our situation into the one described in the proof of Proposition 2.10. Of course, it is important to keep in mind that even though we are phrasing the results in terms of affine plane curves for concreteness, the same discussion applied to projective plane curves without any change.

Remark 2.19. The smoothness of the curve $C/\mathbb{F}_q$ is absolutely crucial for the proof of Proposition 2.10, since the statement is simply not true without this assumption. Consider for example the affine curve $C/\mathbb{F}_q$ given by $X^3 = Y^2$. Let us show that the ring of regular functions at $P = (0,0)$ is not a DVR. Suppose for contradiction $\mathcal{O}_P$ is a DVR and recall that $\mathcal{O}_P$ is the valuation ring of its field of fractions $\mathbb{F}_q(C)$. If we denote the induced valuation by ν_P, we have $\nu_P\left(\frac{Y^2}{X^2}\right) = \nu_p\left(\frac{X^3}{X^2}\right) = \nu_p(X) \geq 1$, since X vanishes at P. Hence, we must have $\nu_P\left(\frac{Y}{X}\right) \geq 1/2$ and so Y/X has positive valuation, hence belonging to the maximal ideal $\mathfrak{M}_P$. Therefore, $Y/X = g/h$, for some polynomials $g, h \in \mathbb{F}_q[X,Y]$ with $h(0,0) \neq 0, g(0,0) = 0$. In other words, we have $X^3 - Y^2 | hY - gX$ and so

$hY = gX + (X^3 - Y^2)q$, for some $q \in \mathbb{F}_q[X,Y]$. Dividing by Y, we obtain $h = X\frac{g+X^2q}{Y} - Yq$, and since h is a polynomial we have $Y|g + X^2q$. But evaluating $h = X\frac{g+X^2q}{Y} - Yq$ at $(0,0)$ shows $h(0,0) = 0$, contradicting the assumption. Hence, $\mathcal{O}_P$ is not a DVR.

Now that we know $\mathcal{O}_P$ is a DVR, we have an induced valuation on the function field of C, denoted by ν_P. If $\nu_P(\varphi) > 0$, we know $\varphi \in \mathfrak{M}_P$ and therefore $\varphi(P)$ is zero. In this case, we say that φ has a zero of order ν_P at P. On the other hand, if $\nu_P(\varphi) < 0$, we say that φ has a pole of order $-\nu_P(\varphi)$ at P.

Earlier, we observed that a DVR R is the set of elements with nonnegative valuation in its field of fractions, under the induced valuation. The same observation applies here, and we see that $\mathcal{O}_P$ is the set of elements $\varphi \in \overline{\mathbb{F}_q}(C)$ with $\nu_P(\varphi) \geq 0$.

Thus far, we have discussed how to associate a valuation to the functions of $\overline{\mathbb{F}_q}(C)$. Since $\mathbb{F}_q(C) \subseteq \overline{\mathbb{F}_q}(C)$, these valuations naturally restrict to the function field $\mathbb{F}_q(C)$. The important property of these restrictions is that if P_1, P_2 are $Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$-conjugate, then ν_{P_1} and ν_{P_2} restrict to the same valuation on $\mathbb{F}_q(C)$. This means that for a prime divisor $P = \{P_1, \dots, P_d\}$, all valuations to its points restrict to the same valuation of $\mathbb{F}_q(C)$ and thus we can talk about the valuation ν_P associated to this divisor.

Proposition 2.20. Let $C/\mathbb{F}_q$ be a projective plane curve and let $P_1, P_2 \in C/\overline{\mathbb{F}_q}$ be points with $\sigma(P_1) = P_2$ for some $\sigma \in Gal(\overline{\mathbb{F}_q}/\mathbb{F}_q)$. Then, the valuations ν_{P_1} and ν_{P_2} agree on $\mathbb{F}_q(C)$.

Proof. The key observation is that for any $g \in \mathbb{F}_q[X,Y,Z]$ we have $g(P_1) = 0$ if and only if $g(P_2) = 0$. This is because applying σ to $g(P_1) = 0$ gives $0 = \sigma g(P_1) = g(\sigma P_1) = g(P_2)$, where g is invariant under σ because $g \in \mathbb{F}_q[X,Y,Z]$. Then, for a rational function $\varphi \in \mathbb{F}_q(C)$, which can be written as $\varphi = \frac{g}{h}$, we have $h(P_1) \neq 0$ if and only if $h(P_2) \neq 0$. This shows that $\nu_{P_1}(\varphi) \geq 0$ is equivalent to $\nu_{P_2}(\varphi) \geq 0$, i.e. $\mathcal{O}_{P_1} \cap \mathbb{F}_q(C) = \mathcal{O}_{P_2} \cap \mathbb{F}_q(C)$. In other words, the valuation rings induced by the restrictions of ν_{P_1}, ν_{P_2} to $\mathbb{F}_q(C)$ are identical, meaning that the induced valuations must be identical too. □

Having defined the valuation associated to a prime divisor P, we also define $\mathcal{O}_P$ to be the set rational functions $\varphi \in \overline{\mathbb{F}_q}(C)$ which are regular at all points of P. Similarly, we define $\mathfrak{M}_P$ to be the set of rational functions vanishing at all points of P. Then, we can describe the residue field of P in the following way.

Proposition 2.21. Let $P = \{P_1, \dots, P_d\}$ be a prime divisor on a curve $C/\mathbb{F}_q$, let $\mathcal{O}_P$ be the ring of regular functions at P and let $\mathfrak{M}_P$ be the maximal ideal of $\mathcal{O}_P$. Then $\dim_{\mathbb{F}_q}(\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C)) = \deg P$.

Proof. The main idea behind this proof is to consider the evaluation map $ev : \mathcal{O}_P \cap \mathbb{F}_q(C) \to \overline{\mathbb{F}_q}$, which evaluates a rational function φ at the point P_1. In other words, we have $ev(\varphi) = \varphi(P_1)$. The kernel of this map is $\mathfrak{M}_P \cap \mathbb{F}_q(C)$ and we claim that its image is $\mathbb{F}_{q^d}$.

By Proposition 1.17, we have that P_1 is a $\mathbb{F}_{q^d}$-rational point of C and therefore P_1 can be written so that all of its coordinates lie in $\mathbb{F}_{q^d}$. This means $\varphi(P_1) \in \mathbb{F}_{q^d}$, showing $\mathrm{im}ev \subseteq \mathbb{F}_{q^d}$. On the other hand, Proposition 1.17 also shows that $\mathbb{F}_{q^d}$ is the smallest extension of $\mathbb{F}_q$ in which the ratios of coordinates of P_1 can lie. Forming the rational functions from the ratios of coordinates of P_1 then show that $\mathrm{im}ev$ contains all of $\mathbb{F}_{q^d}$, completing the proof. □

Proposition 2.22. Let $P_1, \dots, P_n$ be a finite set of distinct prime divisors on a curve $C/\mathbb{F}_q$. There exists a function $\varphi \in \mathbb{F}_q(C)$ such that $\nu_{P_1}(\varphi) = 1$ and $\nu_{P_i}(\varphi) = 0$ for $i \geq 2$.

Proof. For this proof, we assume that we are working with affine curves for the sake of concreteness. Extending it to projective curve makes no big difference. Let $P_1 = \{P_{11}, \dots, P_{1d_1}\}, \dots, P_n = \{P_{n1}, \dots, P_{nd_n}\}$, where $d_i = \deg P_i$. Furthermore, let us assume that P_{ij} has coordinates $P_{ij} = (x_{ij}, y_{ij})$ for all i, j. We will construct φ explicitly in the following way.

Consider the line L in the projective plane $\mathbb{P}^2(\overline{\mathbb{F}_q})$ containing $P_{1,1}$ but no other points $P_{i,j}$. More formally, we pick a polynomial ℓ defining L by setting $g(X,Y) = a(X - x_{11}) + b(Y - y_{11})$ where a, b are chosen such that $g(P_{ij}) \neq 0$. This constraint is equivalent to $a(x_{ij} - x_{11}) + b(y_{ij} - y_{11}) \neq 0$, meaning $[a : b] \neq [y_{ij} - y_{11} : x_{11} - x_{ij}]$. This means that is suffices to choose a, b such that $[a : b]$ avoids a finite set, which we can definitely do since $a, b \in \overline{\mathbb{F}_q}$.

Now, consider the field extension $\mathbb{F}_{q^{d_1}}$, which is the smallest extension of $\mathbb{F}_q$ containing $P_{1,1}, \ldots, P_{1,d_1}$, as guaranteed by Proposition 1.17. Hence, the Galois group, $Gal(\mathbb{F}_{q^d}/\mathbb{F}_q)$, acts transitively and freely on the points of P_1. Consider now the product $h = \prod_{\sigma \in Gal(\mathbb{F}_{q^d}/\mathbb{F}_q)} \sigma g$, where σg is the polynomial obtained by applying σ to all coefficients of g.

Note that σg vanishes at a single point of P_1 and at no other points P_{ij}. Therefore, $\nu_{P_{1j}}(h) = 1$ for all points $P_{1j} \in P_1$. Moreover, h is invariant under $Gal(\mathbb{F}_{q^d}/\mathbb{F}_q)$ action, implying $h \in \mathbb{F}_q[X, Y]$. Hence, h is a $\mathbb{F}_q$-polynomial with $\nu_{P_1}(h) = 1$ and $\nu_{P_i}(h) = 0$ for all $i \geq 2$. □

Remark 2.23. Once we have proven Proposition 2.22, it is very easy to extend it and construct a rational function $\varphi \in \mathbb{F}_q(C)$ with $\nu_{P_i}(\varphi) = k_i$ for any set of integers $k_1, \ldots, k_n$. It suffices to construct $\varphi_1, \ldots, \varphi_n \in \mathbb{F}_q(C)$ having $\nu_{P_j}(\varphi_i) = \delta_{ij}$ and take $\varphi = \prod_{i=1}^{n} \varphi_i^{k_i}$.

2.4 Divisors of rational functions

In this section, we will define the divisors associated to a rational functions.

Definition 2.24. The *divisor* of a nonzero rational function $\varphi \neq 0$ on a curve $C/\mathbb{F}_q$ is defined as

$$\operatorname{div}(\varphi) = \sum_{v_P(\varphi) \neq 0} v_P(\varphi) \cdot P.$$

A priori, it is not obvious that the sum in the definition of $\operatorname{div}(\varphi)$ is finite, and that the above sum is indeed a divisor. It is not hard to see this directly, by showing that two coprime polynomials in the plane intersect only in a finite number of points. For this approach, one may consult Shafarevich [26], page 4. However, we take a slightly different approach, which will allow us to show later that the degree of the divisor defined above is always zero. For this approach, we will show that φ has finitely many zeros, even with multiplicity. Hence, we have

Proposition 2.25. Let φ be a non-constant rational function on the curve $C/\mathbb{F}_q$ and let $\mathbb{F}_q(\varphi)$ be a transcendental extension of $\mathbb{F}_q$ generated by φ. Then, we have

$$\deg\left(\sum_{\nu_P(\varphi) > 0} \nu_P(\varphi) \cdot P\right) \leq \Big[\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)\Big].$$

However, we still need to show that $[\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)] < \infty$. Hence, we begin our argument by showing this first, and then proceeding to prove Proposition 2.25.

Lemma 2.26. If $C/\mathbb{F}_q$ is a smooth projective plane curve given by the equation $F(X, Y, Z) = 0$ and $\varphi \in \mathbb{F}_q(C)$ a nonconstant rational function, then $\mathbb{F}_q(C)/\mathbb{F}_q(\varphi)$ is a finite extension.

Proof. Let $\varphi = g/h$, where $g, h \in \mathbb{F}_q[X, Y, Z]$ are homogeneous polynomial with $d = \deg g = \deg h$. We will show $[\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)] \leq \deg F \cdot d$. To show this, we will show that if $m > \deg F \cdot d$, any m rational functions $\psi_1, \ldots, \psi_m$ satisfy a linear relation over $\mathbb{F}_q[\varphi]$.

To establish some notation, let $\psi_i = \frac{g_i}{h_i}$, where $g_i, h_i \in \mathbb{F}_q[X, Y, Z]$ are homogeneous polynomial with $d_i = \deg g_i = \deg h_i$. The first observation is that a linear relation between $\psi_1, \ldots, \psi_m$ can be rewritten by clearing the denominators in the following way. Starting from

$$\sum_{i=0}^{m} \psi_i \sum_{\ell=0}^{n} a_{k\ell} \varphi^k = 0, \text{ for some } a_{k\ell} \in \mathbb{F}_q,$$

we can clear the denominators to obtain

$$\sum_{k=0}^{m} \left(g_i \prod_{j \neq i} h_j\right) \sum_{\ell=0}^{n} a_{k\ell} g^\ell h_1^{t-\ell} \equiv 0 \pmod{F}.$$

Hence, in order to show that $\psi_1, \dots, \psi_m$ are $\mathbb{F}_q[\varphi]$-linearly dependent, one needs to find a linear relation among polynomials $\left(g_i \prod_{j\neq i} h_j\right) g^\ell h_1^{t-\ell}$, for $i \in \{1, \dots, m\}, \ell \in \{0, \dots, n\}$ modulo F. We will do this using dimension counting.

Namely, let us denote the set of homogeneous polynomials in X, Y, Z of degree $N = \sum_{i=1}^m d_i + td$ by V_N. Note that the polynomials $\left(g_i \prod_{j\neq i} h_j\right) g^\ell h_1^{t-\ell}$ belong to V_N for all $i \in \{1, \dots m\}, \ell \in \{0, \dots, n\}$, and we denote their $\mathbb{F}_q$-span by S_N. Finally, let T_N be the set of polynomials in V_N divisible by F, i.e. $T_N = \{P(x, y, z) \in V_N : F(x, y, z) | P(x, y, z)\}$.

The main goal is to show that S_N and T_N intersect nontrivially, since this gives a nontrivial relation between the polynomials $\left(g_i \prod_{j\neq i} h_j\right) g^\ell h_1^{t-\ell}$ modulo F. To ensure this, we check that $\dim_{\mathbb{F}_q} S_N + \dim_{\mathbb{F}_q} T_N > \dim_{\mathbb{F}_q} V_N$ for large enough n. Let us now compute the dimensions of the relevant spaces.

The space V_N has a basis of monomials $x^i y^j z^k$ with $i + j + k = N$. There are $\binom{N+2}{2}$ such monomials and therefore $\dim V_N = \binom{N+2}{2}$. The dimension of T_N can be computed in a similar way, since this space has the basis consisting of polynomials $x^i y^j z^k F(x, y, z)$, where $i + j + k = N - \deg F$. Therefore, the dimension of the space T_N is equal to the number of triples (i, j, k) of nonnegative integers with $i + j + k = N - \deg F$, which means $\dim_{\mathbb{F}_q} T_N = \binom{N+2-\deg F}{2}$. Finally, the space S_N is a span of $m(n+1)$ polynomials of the form, which are supposed to be linearly independent. Our goal now is to verify that $\dim_{\mathbb{F}_q} S_N > \dim_{\mathbb{F}_q} V_N - \dim_{\mathbb{F}_q} T_N$ for large enough n. This is not hard to see from the following computation

$$\dim_{\mathbb{F}_q} V_N - \dim_{\mathbb{F}_q} T_N = \binom{N+2}{2} - \binom{N+2-\deg F}{2} = nd \deg F + \deg F \sum_{i=1}^m d_i - \frac{\deg F(\deg F - 3)}{2}.$$

On the other hand, we have $\dim_{\mathbb{F}_q} S_N = m(n+1)$, which is linear in n with a leading factor $m > d \deg F$. The conclusion is that for big enough n we have $\dim_{\mathbb{F}_q} S_N > \dim_{\mathbb{F}_q} V_N - \dim_{\mathbb{F}_q} T_N$ just as needed. □

Proof of Proposition 2.25. Before starting the proof, let us observe that φ is a non-constant function and hence not algebraic over $\mathbb{F}_q$. In particular this means that the transcendence degree of $\mathbb{F}_q(\varphi)$ over $\mathbb{F}_q$ is 1 and hence $[\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$ is finite. Hence, this proposition is indeed non-trivial and can be used to show that $\operatorname{div}(\varphi)$ has finite degree.

Now starting the proof, assume for contradiction that $\sum_{\nu_P(\varphi)>0} \nu_P(\varphi) \cdot \deg P > [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$ and pick prime divisors $P_1, \dots, P_l$ with $\sum_{i=1}^l \nu_{P_i}(\varphi) \cdot \deg P_i > [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$.[2]

By Proposition 2.22 we can pick elements t_i for $i = 1, \dots, l$ such that $\nu_{P_j}(t_i) = 0$ for $j \neq i$ and $\nu_{P_i}(t_i) = -1$. Moreover, by Proposition 2.21, for every $i = 1, \dots, l$, we can find a basis $\{u_{im} : m = 1, \dots, \deg P\}$ of $\mathcal{O}_{P_i}/\mathfrak{M}_{P_i}$.

We now claim that the set of rational functions $S = \{u_{im} t_i^{-j} : 1 \leq i \leq l, 1 \leq m \leq \deg P_i, 1 \leq j \leq n_{P_i}\}$ is linearly independent over $k(\varphi)$. Since $u_{im} t_i^{-j}$ are all elements of $\mathbb{F}_q(C)$, we must have $\dim_{\mathbb{F}_q} \mathbb{F}_q(C)/\mathbb{F}_q(\varphi) \geq \sum_{\nu_P(\varphi)>0} \nu_P(\varphi) \cdot \deg P$, which is precisely what we need.

Suppose there was a linear relation among the elements of S, with coefficients $\lambda_{ijm} \in \mathbb{F}_q(\varphi)$. By clearing out the denominators and cancelling the remaining powers of φ, we may assume that all λ_{ijm} are polynomials in φ with coefficients in $\mathbb{F}_q$, and therefore they can be written as $\lambda_{ijm} = \mu_{ijm} + \varphi \kappa_{ijm}$, where $\mu_{ijm} \in \mathbb{F}_q$ and not all μ_{ijm} are equal to zero. Expressed in a formula, we have

$$\sum_{i=1}^{l} \sum_{j=1}^{n_{P_i}} \left(\sum_{m=1}^{\deg P_i} \mu_{ijm} u_{im} \right) t_i^{-j} + \varphi \sum_{i=1}^{l} \sum_{j=1}^{n_{P_i}} \left(\sum_{m=1}^{\deg P_i} \kappa_{ijm} u_{im} \right) t_i^{-j} = 0.$$

Since we know that μ_{ijm} is nonzero for some choice of indices i, j, m, we can take the ν_{P_i} valuation of the linear dependence relation to find:

$$\nu_{P_i} \left(\sum_{i=1}^{l} \sum_{j=1}^{n_{P_i}} \left(\sum_{m=1}^{\deg P_i} \mu_{ijm} u_{im} \right) t_i^{-j} \right) < 0,$$

since the coefficient $\sum_{m=1}^{\deg P_i} \mu_{ijm} u_{im}$ next to t_i^{-j} is nonzero (recall that u_{im} we chosen to be a basis of $\mathcal{O}_P/\mathfrak{M}_P$). On the other hand, we have

[2] If we knew a priori that there were only finitely many prime divisors P where φ vanishes, we could simply pick all of them. However, since we do not know this and since finiteness in used in the proof, we must pick a finite subset where the conclusion is violated (which we can do since $[\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$ is finite).

$$\nu_{P_i}\left(\varphi\sum_{i=1}^{l}\sum_{j=1}^{n_{P_i}}\left(\sum_{m=1}^{\deg P_i}\kappa_{ijm}u_{im}\right)t_i^{-j}\right)=\nu_{P_i}(\varphi)+\nu_{P_i}\left(\sum_{i=1}^{l}\sum_{j=1}^{n_{P_i}}\left(\sum_{m=1}^{\deg P_i}\kappa_{ijm}u_{im}\right)t_i^{-j}\right)\geq\nu_{P_i}(\varphi)-n_{P_i}\geq 0.$$

This presents a contradiction to the linear dependence relation and we conclude that the functions of S are linearly independent. As described above, this completes the proof. □

In light of Proposition 2.25, it is natural to consider the divisor describing the zeros of φ separately from the divisor describing the poles of φ. We can thus make the following definition.

Definition 2.27. Let φ be a rational function on the curve C. Its *zero-divisor* is defined as $\mathrm{div}_0(\varphi)=\sum_{v_P(\varphi)>0}v_P(\varphi)\cdot P$. Its *pole-divisor* is defined as $\mathrm{div}_\infty(\varphi)=\sum_{v_P(\varphi)<0}(-v_P(\varphi))\cdot P$.

A way to rephrase Proposition 2.25 is to say that $\deg(\mathrm{div}_0(\varphi))\leq[\mathbb{F}_q(C):\mathbb{F}_q(\varphi)]$. In other words, Proposition 2.25 states that a rational function φ has only finitely many zeros on the curve C.

Definition 2.28. Let D,E be divisors on the curve $C/\mathbb{F}_q$. We call D and E *linearly equivalent*, or just *equivalent*, if they differ by a principal divisor, i.e. if $D-E=\mathrm{div}(\varphi)$ for some $\varphi\in\mathbb{F}_q(C)$.

We end this section by describing an application of Proposition 2.25 which will be used in Chapter 5.

Proposition 2.29. If $f(X,Y),g(X,Y)\in\mathbb{F}_q[X,Y]$ and $f(X,Y)$ is absolutely irreducible, then $f(X,Y)$ either divides $g(X,Y)$ or has only finitely many roots in common with $g(X,Y)$.

Proof. The idea of the proof is to construct a curve $C/\mathbb{F}_q$ defined by the equation $f(X,Y)=0$ and consider a function $\varphi=\frac{g(X,Y)}{g(X,Y)+1}$ on it. Then, the common roots of f and g correspond to zeros of φ on $C/\mathbb{F}_q$, of which there are only finitely many as shown by Proposition 2.25.[3] □

2.5 Regular functions

The goal of this section is to show that every rational function $\varphi\in\overline{\mathbb{F}_q}(C)$ has zeros and poles on a projective plane curve $C/\overline{\mathbb{F}_q}$. In order to show this, we will introduce a notion of regular functions and characterize regular functions on affine and projective plane curves.

Definition 2.30. A rational function $\varphi\in\overline{\mathbb{F}_q}(C)$ is *regular* if it is regular at every point $P\in C/\overline{\mathbb{F}_q}$. In other words, the set of regular function, denoted by $\mathcal{O}_C$, is defined as $\mathcal{O}_C=\cap_{P\in C/\overline{\mathbb{F}_q}}\mathcal{O}_P$.

Note that $C/\overline{\mathbb{F}_q}$ may be an affine or a projective curve in the above definition. There are several motivations behind introducing this notion, but we will look at it in the context of proving that every nonconstant rational function on a projective curve C has a pole. Namely, if we assume that $\varphi\in\mathbb{F}_q(C)$ has no poles, then φ is a regular function on $C/\overline{\mathbb{F}_q}$. Hence, characterizing regular functions on projective curves directly leads to our goal for this section. However, before we characterize regular functions on projective curves, we will introduce Hilbert's Nullstellensatz and apply it to characterize regular functions on affine curves.

Lemma 2.31 (Weak form of Hilbert's Nullstellensatz)**.** If $I\subset\overline{\mathbb{F}_q}[X_1,\ldots,X_n]$ is an ideal and the polynomials of I have no common zeros, then $I=\overline{\mathbb{F}_q}[X_1,\ldots,X_n]$.

Since the proof of this classical theorem requires a bit of algebraic terminology, we prefer to postpone it to the end of this section, and first show to to apply it to the characterization regular functions.

Proposition 2.32. Let $C/\overline{\mathbb{F}_q}$ be an affine plane curve, given by the equation $f(X,Y)=0$ for some $f\in\mathbb{F}_q[X,Y]$. If $\varphi\in\mathcal{O}_C$ is a regular function on C, then φ is a polynomial, i.e. $\varphi\in K[X,Y]/(f)$.

[3] Technically, we have shown Proposition 2.25 for projective curves, but since restricting to affine curves cannot increase the number of zeros of a rational function, the statement still holds for affine plane curves as well.

Proof. This proof follows is based on the approach given by Shafarevich in [26], page 53. Since φ is regular at every point of $C/\overline{\mathbb{F}_q}$, we have that for any $P \in C/\overline{\mathbb{F}_q}$ the function φ can be written as $\varphi = \frac{g_P}{h_P}$, where $h_P(P) \neq 0$. Consider the ideal generated by h_P for $P \in C/\mathbb{F}_q$ and f, i.e. $I = (\{h_P : P \in C/\mathbb{F}_q\}, f) \subseteq \mathbb{F}_q[X,Y]$. Lemma 2.16 guarantees that $\mathbb{F}_q[X,Y]$ is a Noetherian ring and therefore every ideal of $\mathbb{F}_q[X,Y]$ is finitely generated. This means that there exists a finite collection of points $P_1, \ldots, P_k$ such that $I = (h_{P_1}, \ldots, h_{P_k}, f)$.

Note that the polynomials $h_{P_1}, \ldots, h_{P_k}, f$ do not have a common zero in $\overline{\mathbb{F}_q}^2$. If they had a common zero, say at the point $P \in C/\overline{\mathbb{F}_q}$, we would have $h_P(P) = 0$ since $h_P \in I$, which would contradict the construction of h_P. Therefore, the polynomials in I have no common zero, and Lemma 2.31 guarantees that $I = \mathbb{F}_q[X,Y]$. In particular, there exist polynomials $u_1, \ldots, u_k$ for which $u_1 h_{P_1} + \cdots + u_k h_{P_k} \equiv 1 \pmod f$. Multiplying this relation by φ now gives $\varphi \equiv \sum_{i=1}^k u_i g_i \pmod f$, where both u_i and g_i are in $\mathbb{F}_q[X,Y]$. Hence, φ is a polynomials, as claimed. □

With the characterization of affine regular functions obtained, we can pass to setting of projective plane curves.

Proposition 2.33. Let $C/\overline{\mathbb{F}_q}$ be a projective plane curve. Then every nonconstant rational function $\varphi \in \overline{\mathbb{F}_q}(C)$ has a pole and a zero.

Proof. This proof follows the presentation of Hartshorne, [14], pages 18-19 and Atiyah-MacDonald, [3], page 31. As discussed previously, if φ has no poles on $C/\overline{\mathbb{F}_q}$, then φ is a regular function. Therefore, our aim will be to show that all regular functions are constants. If we show this, the statement easily follows, since any nonconstant φ must have poles, and since φ^{-1} is also nonconstant, it must have poles, which correspond to zeros of φ.

The idea is to consider the restriction of φ onto the subset of $\mathbb{P}^2(\overline{\mathbb{F}_q})$ which corresponds to the affine plane $\overline{\mathbb{F}_q}^2$. More precisely, we define $U = \{[x : y : z] \in \mathbb{P}^2(\overline{\mathbb{F}_q}) : z \neq 0\}$ and consider the affine plane curve $C' = C \cap U$. As discussed in Section 2.1, the curve C' is given by the equation $F(X,Y,1) = 0$, where $F(X,Y,Z) = 0$ is the defining equation of $C/\overline{\mathbb{F}_q}$. Moreover, $\varphi(X,Y,Z) = \frac{G(X,Y,Z)}{H(X,Y,Z)}$ restricts to the function $\varphi'(X,Y) = \varphi(X,Y,1)$. Note that φ' is regular on the affine curve $C'/\overline{\mathbb{F}_q}$ and therefore Proposition 2.32 ensures that φ' is given by a polynomial. In terms of φ, this means that $\varphi(X,Y,1)$ is a polynomial in X, Y and therefore $\varphi(X,Y,Z) = \frac{G_Z(X,Y,Z)}{Z^n Z}$, where $G_Z \in \mathbb{F}_q[X,Y,Z]$ is a homogeneous polynomial of degree N_Z.

The analogous argument show that $\varphi = G_X/X^{N_X}$ and $\varphi = G_Y/Y^{N_Y}$ for some homogeneous polynomials G_X, G_Y of degrees N_X, N_Y. Further, this means that $X^{N_X}\varphi, Y^{N_Y}\varphi, Z^{N_Z}\varphi \in \mathbb{F}_q[X,Y,Z]/(F)$. For simplicity of notation, we will denote the ring $\mathbb{F}_q[X,Y,Z]/(F)$ by R. If we define $N = N_X + N_Y + N_Z$, we see that any monomial $X^a Y^b Z^c$ with $a+b+c = N$ we either have $a \geq N_X, b \geq N_Y$ or $c \geq N_Z$, and therefore $X^a Y^b Z^c \varphi \in R$. In other words, we have $\varphi \mathbb{F}_q{}^{(N)} S_N \subseteq S_N$, where S_N denotes the set of homogeneous polynomials of degree N in $R = \mathbb{F}_q[X,Y,Z]/(F)$. Of course, this implies $\varphi^k S_N \subseteq S_N$ for all $k \geq 0$, and therefore $X^N \varphi^k \in R$ for all $k \geq 0$. This implies $R[\varphi] \subseteq X^{-N}R$, where we consider $X^{-N}R$ to be a finitely generated module over the ring R, which we will denote by M.

Our next step is to show that φ satisfies an algebraic equation over $\overline{\mathbb{F}_q}$ using the analogue of the Cayley-Hamilton for modules. More precisely, if $v_1, \ldots, v_m$ is the generating set of the module M and suppose that $\varphi v_i = \sum_{j=1}^m a_{ij} v_j$ where $a_{ij} \in R$. Then, we have the matrix equation

$$\begin{pmatrix} \varphi - a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & \varphi - a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & \varphi - a_{mm} \end{pmatrix} v_i = 0,$$

for all generators v_i. Let us denote the matrix in the above equation by Φ. Multiplying the above equation with the adjugate of the matrix, we conclude that $\det \Phi \cdot I v_i = 0$ for all v_i. Recall that v_i are elements of $X^{-N}\mathbb{F}_q[X,Y,Z]/(F)$ and therefore the only way we could have $\det \Phi \cdot I v_i = 0$ is that $\det \Phi = 0$. Expanding $\det \Phi = 0$, we obtain the equation of the form $\varphi^m + b_{m-1}\varphi^{m-1} + \cdots + b_0 = 0$, where $b_0, \cdot, b_{m-1} \in \mathbb{F}_q[X,Y,Z]/(F)$. Since φ is homogeneous of degree 0, we may replace b_i by their homogeneous part of degree 0. But these are simply the constant terms in the polynomials b_i and therefore they lie in $\overline{\mathbb{F}_q}$. In other words, we obtain an

equation of the form $\varphi^m + c_{m-1}\varphi^{m-1} + \cdots + c_0 = 0$, where $c_{m-1}, \ldots, c_0 \in \overline{\mathbb{F}_q}$. Since $\overline{\mathbb{F}_q}$ is algebraically closed, any such function φ must be constant, completing the proof. □

The only thing that is left is to prove the weak Nullstellensatz, Lemma 2.31. To prove it, we follow the approach presented by Allcock [1], who follows the original Zariski's proof. We begin by introducing several new algebraic notions.

Definition 2.34. Let L/K be a field extension, and let x be an element of L. We say that x is *algebraic* over K if it satisfies a polynomial relation of the form $x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$ for some elements $a_{n-1}, \ldots, a_0 \in K$. Elements which are not algebraic are called *transcendental*. More generally, we say that elements $x_1, \ldots, x_k \in L$ are *algebraically dependent* if there exists a polynomial $p \in K[X_1, \ldots X_k]$ satisfying $p(x_1, \ldots, x_k) = 0$. Otherwise, we say that these elements are *algebraically independent*. Finally, if all elements of L are algebraic, we say that the extension L/K is *algebraic*.

Proof of Lemma 2.31. Suppose that $I \neq \overline{\mathbb{F}_q}[X_1, \ldots, X_n]$ and let $\mathfrak{m}$ be a maximal ideal containing it. Our goal will be to show that the only maximal ideals of $\overline{\mathbb{F}_q}[X_1, \ldots, X_n]$ are of the form $(X_1 - a_1, \ldots, X_n - a_n)$ for some $a_1, \ldots, a_n \in \overline{\mathbb{F}_q}$. This would imply that all polynomials of I vanish at $(a_1, \ldots, a_n)$ which provides a contradiction.

To prove this statement, we consider the field $L = \overline{\mathbb{F}_q}[X_1, \ldots, X_n]/\mathfrak{m}$, which is a field extension of $\overline{\mathbb{F}_q}$ since $\mathfrak{m}$ is proper. Since $\overline{\mathbb{F}_q}$ is algebraically closed, we have two cases - either $L = \overline{\mathbb{F}_q}$ or L contains a transcendental element over $\overline{\mathbb{F}_q}$.

In the first case, we consider the projection map $\overline{\mathbb{F}_q}[X_1, \ldots, X_n] \to \overline{\mathbb{F}_q}[X_1, \ldots, X_n]/\mathfrak{m} = \overline{\mathbb{F}_q}$ and let a_i be the element of $\overline{\mathbb{F}_q}$ corresponding to the image of X_i under the projection map. This means that $X_i - a_i \in \overline{\mathbb{F}_q}[X_1, \ldots, X_n]$ projects to zero, i.e. that $X_i - a_i \in \mathfrak{m}$. In particular, we see that $(X_1 - a_1, \ldots, X_n - a_n) \subseteq \mathfrak{m}$, but since $(X_1 - a_1, \ldots, X_n - a_n)$ is a maximal ideal, we must have equality, thus proving the statement.

Our goal now is to show that the second case does not occur, i.e. that it is impossible for L to contain transcendental elements over $\overline{\mathbb{F}_q}$. Let $x_1, \ldots, x_m$ be a maximal set of algebraically independent elements over $\overline{\mathbb{F}_q}$.[4] Under this definition, L is an algebraic extension of $\overline{\mathbb{F}_q}(x_1, \ldots, x_m)$. If we define $K = \overline{\mathbb{F}_q}(x_1, \ldots, x_{m-1})$ and write $x = x_m$, then L is an algebraic extension of $K(x)$.

There exists a finite number of elements of L, say $y_1, \ldots, y_m$, such that every other element of L can be written as a $\overline{\mathbb{F}_q}$-polynomial in $y_1, \ldots, y_m$ (for example, one might take the images of $X_1, \ldots, X_n$ in $L = \overline{\mathbb{F}_q}[X_1, \ldots, X_n]/\mathfrak{m}$). In other words, L is a finitely generated $\overline{\mathbb{F}_q}$-algebra. If we pick the minimal such set $y_1, \ldots, y_m$, we can write any element $z \in L$ as

$$z = \sum_{i_1=0}^{d_1} \sum_{i_2=0}^{d_2} \cdots \sum_{i_m=0}^{d_m} \lambda_{i_1,\ldots,i_m} y_1^{i_1} \cdots y_m^{i_m}, \text{ for some } \lambda_{i_1,\ldots,i_m} \in \overline{\mathbb{F}_q}, d_1, \ldots, d_m \in \mathbb{Z}_{\geq 0}. \tag{2.1}$$

Moreover, since $y_1, \ldots, y_m$ are algebraic over $K(x)$, we can choose the degrees $d_1, \ldots, d_m$ in (2.1) to be bounded by the degrees of $y_1, \ldots, y_m$ as algebraic elements over $K(x)$. In other words, every element of L can be written as a linear combination of a finite number of monomials in $y_1, \ldots, y_m$ and thus L is finite-dimensional vector space over $K(x)$.

Let us now pick a linear basis for L over $K(x)$, say $e_1, \ldots, e_d$. Since $y_1, \ldots, y_m$ are certainly linearly independent over $K(x)$, we may assume that $e_1 = y_1, \ldots, e_m = y_m$ and that other elements of the basis are products of y_is. Of course, we can still multiply the elements $e_1, \ldots, e_d$ within the field L and therefore we can write

$$e_i e_j = \sum_{k=1}^{d} \lambda_{ijk} e_k, \text{ for some } \lambda_{ijk} \in K(x).$$

Since $\lambda_{ijk} \in K(x)$, we can write $\lambda_{ijk} = \frac{a_{ijk}(x)}{b_{ijk}(x)}$, for some polynomials $a_{ijk}, b_{ijk} \in K[X]$.

[4] Technically, we need to show that there are no infinite sets of algebraically independent elements. To see this, recall that $L = \overline{\mathbb{F}_q}[X_1, \ldots, X_n]/\mathfrak{m}$, which means that it suffices to show that there are no infinite algebraically independent subsets of $\overline{\mathbb{F}_q}[X_1, \ldots, X_n]$. But it can easily be seen that every $n+1$ polynomials of $\overline{\mathbb{F}_q}[X_1, \ldots, X_n]$ must be algebraically dependent, by a dimension counting argument similar to the one presented in the proof of Lemma 2.26.

Therefore, if we represent $z \in L$ as a linear combination $e_1, \ldots, e_d$ in the form

$$z = \sum_{i_1=0}^{d_1} \sum_{i_2=0}^{d_2} \cdots \sum_{i_m=0}^{d_m} \lambda_{i_1,\ldots,i_m} y_1^{i_1} \cdots y_m^{i_m} = \sum_{i=1}^{k} \frac{a'_k(x)}{b'_k(x)} e_k,$$

where b'_k is a polynomial obtained as a product of $b_{ijk} \in K[X]$. In other words, if we assume that $\frac{a'_k}{b'_k}$ is a reduced fraction, all irreducible factors of b'_k be among irreducible factors of b_{ijk}. But not every element can be expressed in this form - it suffices to pick $z = \frac{1}{x-c} e_1$, where $c \in \overline{\mathbb{F}_q}$ is chosen such that none of the b_{ijk} have $X - c$ as an irreducible factor. This presents a contradiction, showing that the second case is impossible and thus completing the proof. □

Chapter 3

The Riemann-Roch theorem

3.1 Riemann-Roch spaces and the Riemann-Roch theorem

In this section, we will introduce perhaps the most important theorem of the algebraic geometry on curves, the Riemann-Roch theorem. However, before we state the main theorem, let us introduce the fundamental object of interest - the Riemann-Roch spaces.

Definition 3.1. If D is a divisor of $C/\mathbb{F}_q$, the *Riemann-Roch linear space* associated to D is defined as follows

$$\mathcal{L}(D) = \{\varphi \in k(C) : \nu_P(\varphi) + \nu_P(D) \geq 0 \text{ for all } P \in \mathrm{PDiv}(C/\mathbb{F}_q)\} \cup \{0\}.$$

The properties of the valuations ν_P guarantee that $\mathcal{L}(D)$ is a $\mathbb{F}_q$-vector space and therefore we may talk about its dimension. In fact, the Riemann-Roch theorem gives us a formula to compute its dimension.

Theorem 3.2 (Riemann-Roch)**.** Let $C/\mathbb{F}_q$ be a projective plane curve. There exists an integer g, called the *genus* of the curve C, and a divisor K of degree $2g-2$, called the *canonical divisor*, such that the following relation is true for any divisor D:

$$\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g + 1 + \dim_{\mathbb{F}_q} \mathcal{L}(K - D).$$

Given the importance of this result, before trying to prove it we will overview on the high level one possible interpretation, as well as general steps in the proof. One should think of the right hand side of the equation as having three main terms $\deg D + 1$, $-g$ and $\dim \mathcal{L}(K - D)$. We will begin the proof of Riemann-Roch theorem by showing that $\dim \mathcal{L}(D) \leq \deg D + 1$. In fact, in the proof we will use only the local behavior of the rational functions, which will make the whole proof much easier. Furthermore, Riemann's theorem shows that the actual dimension is still within a constant of the value predicted by local considerations. In other words, Riemann's theorem states that $\dim_{\mathbb{F}_q} \mathcal{L}(D) \geq \deg D + 1 - g$, for some finite integer g depending only on the curve $C/\mathbb{F}_q$, which is called the genus. Finally, pinning down the exact value of $\dim_{\mathbb{F}_q} \mathcal{L}(D)$ between $\deg D + 1 - g$ and $\deg D + 1$ will be the hardest part of the proof, and it will require us to introduce several new objects, such as adèles and differentials.

Before launching into the proof, we begin with preliminary lemmas which will help set the ground for the proof in the later sections.

Proposition 3.3. Let $D \leq E$ be divisors of $C/\mathbb{F}_q$. Then, we have $\mathcal{L}(D) \subseteq \mathcal{L}(E)$ and

$$\dim_{\mathbb{F}_q} \mathcal{L}(E)/\mathcal{L}(D) \leq \deg E - \deg D.$$

Proof. The first statement, that $\mathcal{L}(D) \subseteq \mathcal{L}(E)$, follows almost from the definition and we will focus on showing only the second statement. Since $E - D = \sum_P n_P \cdot P$ for positive values of n_P, we will prove the statement by induction on $\sum_P n_P$, where the base case is trivial since $\dim \mathcal{L}(D)/L(D) = 0$.

For the induction step, we essentially need to show that the statement holds when $E = D + P$ for a prime divisor $P = \{P_1, \ldots, P_{\deg P}\}$. We will show $\dim \mathcal{L}(E)/\mathcal{L}(D) \leq \deg P$. To do this, we fix an arbitrary element

$\varphi \in \mathbb{F}_q(C)$ with $\nu_P(\varphi) = \nu_P(E)$. This is possible, since it suffices to choose any element $\varphi \in \mathfrak{M}_P^{\nu_P(E)} \setminus \mathfrak{M}_P^{\nu_P(E)+1}$ (and this difference is not empty since ν_P is surjective).

Consider now a linear transformation $T : \mathcal{L}(E) \mapsto \mathcal{O}_P \cap \mathbb{F}_q(C)$ defined by $\psi \mapsto \psi\varphi$. Note that $\psi\varphi$ is indeed regular at P, since the assumption implies $\nu_P(\varphi\psi) = \nu_P(\varphi) + \nu_P(\psi) \geq \nu_P(E) - \nu_P(E) \geq 0$. Furthermore, we have $\mathcal{L}(D) = \ker T$. The reason for this is simple - for $\psi \in \mathcal{L}(D)$ we have $\nu_P(\psi\varphi) = \nu_P(\psi) + \nu_P(\varphi)$ and thus $\psi\varphi$ vanishes at P if and only if $\nu_P(\psi) \geq -\nu_P(\varphi) + 1 = \nu_P(D)$. Therefore, T induces a injective linear map $T : \mathcal{L}(E)/\mathcal{L}(D) \to (\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C))$, showing that $\dim_{\mathbb{F}_q} \mathcal{L}(D)/\mathcal{L}(E) \leq \dim_{\mathbb{F}_q}(\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C))$. Since $\dim_{\mathbb{F}_q}(\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C)) = \deg P$ by Proposition 2.21, we obtain $\dim_{\mathbb{F}_q} \mathcal{L}(E)/\mathcal{L}(D) \leq \deg P$, as claimed. □

Corollary 3.4. The space $\mathcal{L}(D)$ is a finite-dimensional vector space over $\mathbb{F}_q$. Furthermore, if D is positive we have $\dim_{\mathbb{F}_q} \mathcal{L}(D) \leq \deg D + 1$.

Proof. It suffices to consider only positive divisors, since any for any divisor D one can find a positive divisor E with $D \leq E$ and $\mathcal{L}(D) \subset \mathcal{L}(E)$.

Every positive divisor can be written as a sum of prime divisors $D = \sum_P n_P P$. We will show the statement that $\dim_{\mathbb{F}_q} \mathcal{L}(D) \geq \deg D + 1$ by induction on $\sum_P n_P$. In the base case, $D = 0$, we have $\mathcal{L}(0) = \mathbb{F}_q$, since every non-constant function has a pole on $C/\mathbb{F}_q$ by Proposition 2.33. Hence, the dimension of $\mathcal{L}(0)$ is 1.

To perform an induction step, we consider $E = D + P$, where P is a prime divisor recall that Proposition 3.3 shows that $\dim_{\mathbb{F}_q} \mathcal{L}(E)/\mathcal{L}(D) \leq \deg P$. Combining this with the induction hypothesis which states $\dim_{\mathbb{F}_q} \mathcal{L}(D) \leq \deg D + 1$, we obtain $\dim_{\mathbb{F}_q} \mathcal{L}(E) \leq \deg D + \deg P + 1 = \deg E + 1$, as needed. □

Here is another basic tool we will use to understand the Riemann-Roch spaces, which shows that equivalent divisors have isomorphic Riemann-Roch spaces. More precisely, we have the following isomorphism of vector spaces.

Proposition 3.5. If $D \sim E$ are equivalent divisors of $C/\mathbb{F}_q$, in the sense that $D + \operatorname{div}(\varphi) = E$ for a nonzero rational function $\varphi \in \mathbb{F}_q(C)$, then $\mathcal{L}(D) \cong \mathcal{L}(E)$.

Proof. Consider the isomorphism $T : \mathcal{L}(D) \to \mathcal{L}(E)$ given by $\psi \mapsto \psi\varphi$. The conditions $\nu_P(\psi) \geq -\nu_P(D)$ are equivalent to $\nu_P(\psi\varphi) \geq -\nu_P(D) - \nu_P(\varphi) = -\nu_P(E)$, and hence we have $\psi\varphi \in \mathcal{L}(E)$. Checking that T is a homomorphism is direct and since T has an inverse given by $\vartheta \mapsto \vartheta\varphi^{-1}$ it is also an isomorphism. □

We will now use the Riemann-Roch spaces and Corollary 3.4 to show a general fact about rational functions on a curve, that every function have an equal number of zeros and poles when counted with multiplicity. To show this, we begin by showing the equality holds in Proposition 2.25.

3.2 Interlude: Rational functions have equally many zeros and poles

In this section, we will divert our attention for a second from the aim of proving the Riemann-Roch theorem and demonstrate how the Riemann-Roch spaces that were just introduced can help us show that a rational function has equally many zeros and poles, when counted with multiplicity.

Proposition 3.6. Let φ be a rational function on $C/\mathbb{F}_q$. Then $\deg(\operatorname{div}_0(\varphi)) = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$

Proof. The upper bound on $\deg(\operatorname{div}_0(\varphi))$ was already shown in Proposition 2.25. Now, we focus on showing that $\deg(\operatorname{div}_0(\varphi)) \geq [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$. The overall idea of the proof is the following. We can use Corollary 3.4 to obtain a lower bound on the degree of a divisor in terms of $\dim_{\mathbb{F}_q} \mathcal{L}(D)$, for some divisor D. Then, the goal would be to relate the $\mathbb{F}_q$-dimension of $\mathcal{L}(D)$ and $\mathbb{F}_q(\varphi)$-dimension of $\mathbb{F}_q(C)$ by transforming a basis of the $\mathbb{F}_q(C)$ into a large independent set in $\mathcal{L}(D)$. Now, we will present the details of the argument, starting from the way to relate the dimensions of the mentioned vector spaces.

Let $\psi_1, \ldots, \psi_n$ be the $\mathbb{F}_q(\varphi)$-basis of $\mathbb{F}_q(C)$, where $n = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$, and let $D = \sum_{i=1}^n \operatorname{div}_\infty(\psi)$. Then, we know that $\psi_i \in \mathcal{L}(D)$, and moreover for any $l \geq 0$ we have $\varphi^{-l}\psi_i \in \mathcal{L}(D + l\operatorname{div}_0(\varphi))$. In other words, if we fix an integer m, the space $\mathcal{L}(D + m\operatorname{div}_0(\varphi))$ contains a set of elements $S = \{\varphi^{-l}\psi_i : 0 \leq l \leq m, 1 \leq i \leq n\}$. Since

the functions ψ_i are linearly independent over $\mathbb{F}_q(\varphi)$, we conclude that the functions of the set S are linearly independent over $\mathbb{F}_q$. Hence, $\dim_{\mathbb{F}_q} \mathcal{L}(D + m\mathrm{div}_0(\varphi)) \geq n(m+1)$.

On the other hand, Corollary 3.4 implies that $\dim_{\mathbb{F}_q} \mathcal{L}(D + m\mathrm{div}_0(\varphi)) \leq m\deg(\mathrm{div}_0(\varphi)) + \deg D + 1$. Combining the two derived inequalities, we obtain $m(\deg(\mathrm{div}_0(\varphi) - n) \geq n - \deg(D) - 1$. Since this holds for any m, we must have $\deg(\mathrm{div}_0(\varphi)) \geq n = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$, which completes the proof. □

Corollary 3.7. Let $\varphi \in k(C)$ be a nonzero rational function. Then $\deg \mathrm{div}(\varphi) = 0$. In other words, all principal divisors have degree 0.

Proof. We will assume that φ is not a constant function, since the statement is otherwise trivial. From the definition of $\mathrm{div}_0(\varphi), \mathrm{div}_\infty(\varphi)$ we have $\mathrm{div}(\varphi) = \mathrm{div}_0(\varphi) - \mathrm{div}_\infty(\varphi)$ and hence $\deg \mathrm{div}(\varphi) = \deg \mathrm{div}_0(\varphi) - \deg \mathrm{div}_\infty(\varphi)$. Proposition 3.6 shows that $\deg \mathrm{div}_0(\varphi) = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$, while applying the same proposition to φ^{-1} shows $\deg \mathrm{div}_\infty \varphi = \deg \mathrm{div}_0 \varphi^{-1} = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi^{-1})] = [\mathbb{F}_q(C) : \mathbb{F}_q(\varphi)]$. Hence, the degrees of $\mathrm{div}_0\varphi$ and $\mathrm{div}_\infty\varphi$ are equal, completing the proof. □

A direct consequence of Corollary 3.7 is that equivalent divisors have the same degree. Another corollary of this result allows us to compute $\mathcal{L}(D)$ for all divisors D of negative degree.

Corollary 3.8. If D is a divisor of $C/\mathbb{F}_q$ with $\deg D < 0$, then $\mathcal{L}(D) = \{0\}$.

Proof. For any nonzero rational function $\varphi \in \mathbb{F}_q(C)$ we have $\deg \mathrm{div}(\varphi) = 0$, and thereofre $\deg(\mathrm{div}(\varphi) + D) < 0$. Hence, $\deg \varphi + D$ cannot be a positive divisor. □

3.3 Riemann's inequality

We are now ready to prove the first part of the Riemann-Roch theorem, Riemann's inequality.

Proposition 3.9. There exists an integer $g \geq 0$ depending on the curve $C/\mathbb{F}_q$ such that for all divisors D one has $\dim_{\mathbb{F}_q} \mathcal{L}(D) \geq \deg(D) + 1 - g$.

One way to interpret Riemann's theorem is that it shows tightness of inequalities given in Corollary 3.4 up to an additive error of g. However, this is only half of the way to the full Riemann-Roch theorem which pins down the exact value of the error term between $\dim_{\mathbb{F}_q} \mathcal{L}(D)$ and $\deg D + 1$.

Proof of Proposition 3.9. For a general divisor D, we denote by $s(D)$ the difference $(\deg D + 1) - \dim_{\mathbb{F}_q} \mathcal{L}(D)$. To goal is to find an integer g such that $g \geq s(D)$ for all D.

Let us begin by picking an arbitrary function $\varphi \in \mathbb{F}_q(C)$ and repeating the procedure in the proof of Proposition 3.6 to find a divisor A for which $\dim_{\mathbb{F}_q} \mathcal{L}(m\mathrm{div}_0(\varphi) + A) \geq \deg(\mathrm{div}_0(\varphi))(m+1)$. Denoting $\mathrm{div}_0(\varphi) = B$ for simplicity of notation, using Proposition 3.3 we find that

$$\dim_{\mathbb{F}_q} \mathcal{L}(mB) + \deg A \geq \dim \mathcal{L}(mB + A) \geq (m+1)\deg B.$$

Therefore, $s(mB) = \deg A - \deg B + 1$, which is independent of m. Hence, we conclude that there exists an integer g for which $s(mB) \leq g$ for all m.

Before continuing the proof, let us add a motivational paragraph explaining the intuition behind the rest of the proof. The only remaining thing to do is to extend the bound $s(D) \leq g$ to all divisors D and not only to those of the form $D = mB$ for some m. Let us fix a general divisor D, with the aim of showing $s(D) \leq g$ for the integer g defined as above. We use a simple trick - if $D \leq E$ are divisors, Proposition 3.3 directly implies $s(D) \leq s(E)$. If we could find a divisor E for which $E \leq mB$ for some m, we would obtain $s(D) \leq s(E) \leq s(mB) \leq g$, completing the proof. Unfortunately this is not quite possible. Instead, we will try to find a divisor $F \sim E$ for which $F \leq mB$. The reason this is useful is because $E \sim F$ implies $\deg E = \deg F$ and $\dim \mathcal{L}(E) = \dim \mathcal{L}(F)$, implying $s(E) = s(F)$. Then, we will have the chain of inequalities $s(D) \leq s(E) = s(F) \leq s(mB) \leq g$, and thus complete the proof.

Let us now present the details for the remainder of the proof. Having fixed an arbitrary divisor D, we pick an arbitrary positive divisor $E \geq D$ and note that

$$\dim \mathcal{L}(mB - E) \geq \dim \mathcal{L}(mB) - \deg E \geq \deg mB + 1 - g - \deg E.$$

Hence, for big enough m, one can ensure that the space $\mathcal{L}(mB - E)$ has dimension ≥ 2 and thus contains a non-constant function φ. Choosing $F = E - \mathrm{div}\varphi$, we find that $F \leq mB$ from the assumption $\varphi \in \mathcal{L}(mB - E)$. Hence, we obtain our chain of inequalities $s(D) \leq s(E) = s(F) \leq s(mB) \leq g$, completing the proof. □

We will use Proposition 3.9 to define the genus of the curve.

Definition 3.10. The smallest integer g for which Riemann's inequality is satisfied is called the *genus* of the curve $C/\mathbb{F}_q$. In other words, $g = \max_{D \in \mathrm{Div}(C/\mathbb{F}_q)} \deg D + 1 - \dim_{\mathbb{F}_q} \mathcal{L}(D)$.

Even though this definition is sufficient for our purposes, it is not very convinient to work with. For example, from this definition, it is a priori whether the genus of $C/\mathbb{F}_q$ is the same as the genus of $C/\mathbb{F}_{q^m}$.

Let us now show that, with the genus defined as above, Riemann's inequality is tight for all divisors of large enough degree.

Proposition 3.11. Let D be a divisor of $C/\mathbb{F}_q$ of large enough degree, depending on $C/\mathbb{F}_q$. Then, $\dim \mathcal{L}(D) = \deg D + 1 - g$, where g is the genus of the curve.

Proof. By the definition of g, we have a divisor $E \in \mathrm{Div}(C/\mathbb{F}_q)$ for which $g = \deg E + 1 - \dim_{\mathbb{F}_q} \mathcal{L}(E)$. We claim that if $\deg D \geq \deg E + g$, one must have $\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g$. There are two main steps towards obtaining this, both of which were already shown in esentially the same form in the proof of Riemann's inequality. The first step will be to show that for all divisors $F \geq E$ we have $\dim_{\mathbb{F}_q} \mathcal{L}(F) = \deg F + 1 - g$. The second step will then be to show that D is linearly equivalent to a divisor $F \geq E$.

To show the first step, take $F \geq E$ and recall that we have shown $s(F) \geq s(E) = g$ in the proof of Riemann's theorem, while we know that $s(F) \leq g$ for all divisors F. Hence, $s(F) = g$, or in other words $\dim_{\mathbb{F}_q} \mathcal{L}(F) = \deg F + 1 - g$. As for the second step, our goal is to find a rational function φ for which $D + \mathrm{div}(\varphi) \geq E$, i.e. $\mathrm{div}(\varphi) + (D - E) \geq 0$. Hence, it suffices to show that $\dim_{\mathbb{F}_q} \mathcal{L}(D - E) \geq 1$, which is true by Riemann's inequality since

$$\dim_{\mathbb{F}_q} \mathcal{L}(D - E) \geq \deg(D - E) + 1 - g \geq 1.$$

Hence, both steps of our plan combine to give $s(D) = s(D + \mathrm{div}(\varphi)) = g$, meaning $\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g$. □

3.4 Adèles and differentials

In order to pin down the exact error term in the Riemann-Roch theorem, we introduce several more abstract notions, including the notion of adèles and differentials. Before diving into the definitions, let us say a couple of words about the reasons for introducing them.

When looking at rational functions on a given algebraic curve, one might consider their local properties, such as the order of vanishing at a particular prime divisor, or global properties, such as having a zero or a pole. The general intuition we have obtained so far is that the local properties are comparatively easy to understand. For example, Proposition 2.22 allows us to construct functions satisfying arbitrary local properties. However, when viewed from the global perspective, the rational functions are much more rigid - they must have the same number of zeros and poles, they satisfy the continuation principle and so on.

This interplay allows us to understand why the dimension of $\mathcal{L}(D)$ is not simply $\deg D + 1$, as one would expect from local considerations. The reason for this is simple enough - not every function with prescribed zeros and poles can be extended to the whole curve without introducing any new poles. In a way, the global conditions on the rational functions is what makes the understanding the size of $\mathcal{L}(D)$ an intricate business. Hence, in an attempt to understand the interplay between global and local conditions better, we will introduce a local analogue of rational functions, which are not bound to satisfy any rigid global constraints, and we will call them adèles.

Definition 3.12. An *adèle* of the curve $C/\mathbb{F}_q$ is a map $\alpha : \mathrm{PDiv}(C/\mathbb{F}_q) \to \mathbb{F}_q(C)$ which assigns to every prime divisor $P \in \mathrm{PDiv}(C/\mathbb{F}_q)$ of $C/\mathbb{F}_q$ a rational function $\alpha(P) \in \mathbb{F}_q(C)$ such that $\nu_P(\alpha(P)) < 0$ for only finitely many prime divisors P.

As we suggested previously, the adèles are objects which resemble functions around every single point, but which are not bound to satisfy any rigid global constraints. Thus, our program for understanding the dimension of $\mathcal{L}(D)$ from now on will have two steps. The first step will be to generalize the notions of Riemann-Roch spaces to adèles and compute the dimension of the corresponding spaces. Then, our second step will be to relate the adèle spaces to the usual Riemann-Roch spaces and compute the dimension of $\mathcal{L}(D)$ through the use of differentials.

Let us begin slowly by carrying over the notions defined for functions to the world of adèles. For the curve $C/\mathbb{F}_q$, we denote the space of all adèles by $\mathcal{A}$. The space $\mathcal{A}$ can be made into a $\mathbb{F}_q$-vector space by defining coordinatewise addition and multiplication. Furthermore, one can extend the valuations ν_P to $\mathcal{A}$ by defining $\nu_P(\alpha) = \nu_P(\alpha(P))$. In analogy with the definition of the Riemann-Roch spaces, for a divisor $D \in \operatorname{Div}(C/\mathbb{F}_q)$, we can define $\mathcal{A}(D) = \{\alpha \in \mathcal{A} : \nu_P(\alpha) + \nu_P(D) \geq 0 \text{ for all } P\}$. Finally, every rational function $\varphi \in \mathbb{F}_q(C)$ has an adèle α_φ naturally associated to it which satisfies $\alpha_\varphi(P) = \varphi$ for all P. Hence, from now on we will abuse the terminology slightly by thinking of $\mathbb{F}_q(C)$ as a subspace of $\mathcal{A}$ and calling its elements *principal adèles*.

Having introduces $\mathcal{A}$ as a vector space over $\mathbb{F}_q$, a word of caution is needed. It is quite clear that $\mathcal{A}$ is infinite-dimensional, precisely because no rigid global constrains are imposed, and hence we need to work with quotients in order to compute various dimensions. However, this does not present a major problem, and we have our first proposition, analogous to Proposition 3.3.

Proposition 3.13. Let $D \leq E$ be divisors of $C/\mathbb{F}_q$. Then $\mathcal{A}(D) \subset \mathcal{A}(E)$ and $\dim_{\mathbb{F}_q} \mathcal{A}(E)/\mathcal{A}(D) = \deg E - \deg D$.

Proof. This proof completely mirrors the proof of Proposition 3.3. It suffices to consider the case $E = D+P$, for a prime divisor P, and so we pick $\varphi \in \mathbb{F}_q(C)$ such that $\nu_P(\varphi) = \nu_P(E)$. Then, we construct a isomorphism of vector spaces $T : \mathcal{A}(E)/\mathcal{A}(D) \to (\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C))$ induced by $\alpha \mapsto \varphi\alpha(P)$, where α_φ denotes the principal adèle corresponding to φ. The map is is injective, as in the case $\mathcal{L}(E)/\mathcal{L}(D) \to (\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C))$, but it is also surjective. To see this, pick an arbitrary element $\psi \in \mathcal{O}_P$ and define an adèle α by setting $\alpha(P) = \varphi^{-1}\psi$ and $\alpha(Q) = 0$ for $Q \neq P$. Then we must have $T(\alpha) = \varphi(\varphi^{-1}\psi) = \psi$, as needed.

Hence, we have $\dim_{\mathbb{F}_q} \mathcal{A}(E)/\mathcal{A}(D) = \dim_{\mathbb{F}_q}(\mathcal{O}_P \cap \mathbb{F}_q(C))/(\mathfrak{M}_P \cap \mathbb{F}_q(C)) = \deg P$, by Proposition 2.21. □

Now, we will turn to the connection between the spaces $\mathcal{A}(D)$ and $\mathcal{L}(D)$. To this end, fix divisors $D \leq E$ and introduce the spaces $\mathcal{A}(D) + \mathbb{F}_q(C)$ which are simply the direct sum of subspaces corresponding to $\mathcal{A}(D)$ and the principal adèles.

Consider now the map $\sigma : \mathbb{F}_q(C) \to \mathcal{A}$ given by $\sigma(\varphi) = \alpha_\varphi$ and note that $\nu_P(\varphi) = \nu_P(\sigma(\varphi))$ by definition. Hence, σ also takes $\mathcal{L}(D)$ to $\mathcal{A}(D)$, and hence we have the following diagram.

$$\begin{array}{ccccc} \mathcal{L}(E) & \xrightarrow{\sigma} & \mathcal{A}(E) & \xrightarrow{\pi} & \mathcal{A}(E)/\mathcal{A}(D) \\ \uparrow{\scriptstyle \iota} & & \uparrow{\scriptstyle \iota} & & \\ \mathcal{L}(D) & \xrightarrow{\sigma} & \mathcal{A}(D) & & \end{array}$$

Noting that $\mathcal{L}(D)$ lies in the kernel of the composition $\pi \circ \sigma \circ \iota$, we see that $\pi \circ \sigma$ can be extended to a map $\mathcal{L}(E)/\mathcal{L}(D) \to \mathcal{A}(E)/\mathcal{A}(D)$. By abuse of notation, we will call the induced map σ too.

In the exact same way, one can construct the map $\tau : \mathcal{A}(E)/\mathcal{A}(D) \mapsto (\mathcal{A}(E) + \mathbb{F}_q(C))/(\mathcal{A}(D) + \mathbb{F}_q(C))$ given by $\tau(\alpha + \mathcal{A}(D)) = \alpha + \mathcal{A}(D) + \mathbb{F}_q(C)$. Now, we have a proposition that establishes the connection between the Riemann-Roch spaces and adèles.

Proposition 3.14. We have the following short exact sequence:

$$0 \mapsto \frac{\mathcal{L}(E)}{\mathcal{L}(D)} \xrightarrow{\sigma} \frac{\mathcal{A}(E)}{\mathcal{A}(D)} \xrightarrow{\tau} \frac{\mathcal{A}(E) + \mathbb{F}_q(C)}{\mathcal{A}(D) + \mathbb{F}_q(C)} \mapsto 0.$$

As a consequence, we have $\dim_{\mathbb{F}_q} \mathcal{A}(E)/\mathcal{A}(D) = (\deg E - \dim_{\mathbb{F}_q} \mathcal{L}(E)) - (\deg D - \dim_{\mathbb{F}_q} \mathcal{L}(D))$.

Proof. We need to check three conditions, none of which are particularly hard. Namely, we need to show that σ is injective, that τ is surjective and that $\operatorname{im}\sigma = \ker\tau$.

The injectivity of σ follows easily, since $\alpha_\varphi \in \mathcal{A}(D)$ means precisely that $\nu_P(\alpha_\varphi) = \nu_P(\alpha_\varphi(P)) = \nu_P(\varphi) \geq \nu_P(-D)$ and so $\varphi \in \mathcal{L}(D)$. Showing surjectivity of τ is also simple since we have $\tau(\alpha+\mathcal{A}(D)) = \alpha+\mathcal{A}(D)+\mathbb{F}_q(C)$ for any $\alpha \in \mathcal{A}(E)$. Finally, showing that $\operatorname{im}\sigma = \ker\tau$ is a little harder.

First of all, we have $\operatorname{im}\sigma \subseteq \ker\tau$ since $\alpha_\varphi \in \mathbb{F}_q(C)$ for all φ. On the other hand, if $\tau(\alpha) = 0$ for some $\alpha \in \mathcal{A}(E)$, we must have $\alpha + \mathcal{A}(D) = \alpha_\varphi + \mathcal{A}(D)$ for some $\alpha_\varphi \in \mathbb{F}_q(C)$. Then, we have $\alpha - \alpha_\varphi \in \mathcal{A}(D) \subset \mathcal{A}(E)$ and therefore $\alpha_\varphi \in \mathcal{A}(E)$. This also means that $\varphi \in \mathcal{L}(E)$ and hence $\sigma(\varphi + \mathcal{L}(D)) = \alpha_\varphi + \mathcal{A}(D) = \alpha + \mathcal{A}(D)$. This shows $\ker\tau \subseteq \operatorname{im}\sigma$ and completes the proof. □

Let us now introduce the last missing ingredient for the proof of the Riemann-Roch theorem, the notion of the differential.

Definition 3.15. A *differential* ω on a curve $C/\mathbb{F}_q$ is simply a linear map $\omega : \mathcal{A} \to \mathbb{F}_q$ which vanishes on $\mathcal{A}(D) + \mathbb{F}_q(C)$ for some divisor $D \in \operatorname{Div}(C/\mathbb{F}_q)$.

Remark 3.16. Although this definition may seem unintuitive, it mirrors the notion of the differential in the theory of Riemann surfaces very closely. Namely, a meromorphic differential form ω on a Riemann surface X can be locally expressed as in the form $f\ dz$ where f is a meromorphic function and z a local coordinate. Thus, the differential form may be integrated against any meromorphic function h on a Riemann surface, thus producing a pairing $(h, \omega) \mapsto \int_X h\omega$. Furthermore, if one prescribes a zero of high enough order for h, i.e. we assume $\operatorname{div}_0(h) \geq \operatorname{div}_\infty(\omega)$ at every point where ω has a pole, the resulting product will have no poles, and hence the integral $\int_X h\omega$ would be equal to zero. However, it is not possible to find such a function h because of the "global constraints", but it is possible in the case of adèles.

Let Ω be the set of all differentials ω and let $\Omega(D)$ be the set of differentials vanishing on $\mathcal{A}(D) + \mathbb{F}_q(C)$. The set $\Omega(D)$ is clearly a $\mathbb{F}_q$-vector space and we will now see that the space $\Omega(D)$ corresponds exactly to the error term in the Riemann-Roch theorem.

Proposition 3.17. Let D be a divisor on the curve $C/\mathbb{F}_q$. Then

$$\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g + \dim_{\mathbb{F}_q} \Omega(D).$$

Remark 3.18. Note that Proposition 3.17 almost completes the proof of the Riemann-Roch theorem, and the only remaining ingredient will be showing that $\dim_{\mathbb{F}_q} \Omega(D) = \dim_{\mathbb{F}_q} \mathcal{L}(K - D)$ for an appropriately defined divisor K. This will be the main goal of the Section 3.5.

Proof. Let us rewrite the inequality in the following form $\dim_{\mathbb{F}_q} \mathcal{L}(D) - \deg D - 1 + g = \dim_{\mathbb{F}_q} \Omega(D)$. Let $E \geq D$ be a divisor of high enough degree, such that $s(E) = g$. The proof will then proceed in three steps, following this chain of equalities

$$\dim_{\mathbb{F}_q} \mathcal{L}(D) - \deg D - 1 + g = \dim_{\mathbb{F}_q} \frac{\mathcal{A}(E) + \mathbb{F}_q(C)}{\mathcal{A}(D) + \mathbb{F}_q(C)} = \dim_{\mathbb{F}_q} \frac{\mathcal{A}}{\mathcal{A}(D) + \mathbb{F}_q(C)} = \dim_{\mathbb{F}_q} \Omega(D).$$

Under the above choice of E, the first step is a direct application of Proposition 3.14. To show the second inequality, it suffices to show that $\mathcal{A}(F) + \mathbb{F}_q(C) = \mathcal{A}(E) + \mathbb{F}_q(C)$ for all $F \geq E$, since $\mathcal{A} = \bigcup_{F \in \operatorname{Div}(C/\mathbb{F}_q)} \mathcal{A}(F)$. Recall from the proof of Proposition 3.11 that $s(F) = s(E) = g$, and therefore Proposition 3.14 shows that $\dim_{\mathbb{F}_q}(\mathcal{A}(F) + \mathbb{F}_q(C))/(\mathcal{A}(E) + \mathbb{F}_q(C)) = (s(F) - 1) - (s(E) - 1) = 0$. Finally, to show the third step, recall that all differentials of $\Omega(D)$ vanish identically on $\mathcal{A}(D) + \mathbb{F}_q(C)$, and therefore are in bijection with the linear maps from $\mathcal{A}/(\mathcal{A}(D) + \mathbb{F}_q(C))$ to $\mathbb{F}_q$. Since basic linear algebra guarantees that the dimension of the space and its dual is the same, the proof is complete. □

3.5 Serre duality

The main goal of this section will be to relate the dimensions of the spaces $\mathcal{L}(K - D)$ and $\Omega(D)$, for some divisor K. Before we can even formulate this statement precisely, we need show how to define the divisor K in question. We will do this by showing how to associate divisors to differentials.

Proposition 3.19. Let ω be a nonzero differential on the curve $C/\mathbb{F}_q$ and let S be the set of divisors $D \in \mathrm{Div}(C/\mathbb{F}_q)$ for which ω vanishes identically on $\mathcal{A}(D) + \mathbb{F}_q(C)$. Then S contains a maximal element with respect to the relation $\leq$, i.e. there is a divisor $K \in S$ such that $D \leq K$ for all $D \in S$.

Proof. The proof of this lemma is just a infinite-dimensional version of the fact that every linear map has a maximal subspace on which it vanishes, called the kernel. We rely on two simple claims: that if $D, E \in S$, then $\max\{D, E\} \in S$[1] as well and that divisors in S have bounded degree. This shows that any ascending chains in S are finite, which suffices to show that S has a maximal element.

To show the first step, pick arbitrary divisors $D, E \in S$ and note that since ω vanishes on $\mathcal{A}(D) + \mathbb{F}_q(C)$, $\mathcal{A}(E) + \mathbb{F}_q(C)$, it must vanish on $\mathcal{A}(D) + \mathcal{A}(E) + \mathbb{F}_q(C)$ too. Hence, it is sufficient to prove that $\mathcal{A}(F) \subset \mathcal{A}(D) + \mathcal{A}(E)$, since this shows that $\mathcal{A}(F) + \mathbb{F}_q(C) \subset \mathcal{A}(D) + \mathcal{A}(E) + \mathbb{F}_q(C)$ and thus $\omega|_{\mathcal{A}(F)+\mathbb{F}_q(C)} \equiv 0$. Let $\alpha \in \mathcal{A}(F)$ and let $T = \{P \in \mathrm{PDiv}(C/\mathbb{F}_q) : \nu_P(\alpha) + \nu_P(D) < 0\}$. Note that at all prime divisors $P \in T$ we have $\nu_P(\alpha) \geq -\nu_P(F)$ and hence $\nu_P(\alpha) \geq -\nu_P(E)$. Therefore, one may write $\alpha = \alpha_D + \alpha_E$, where

$$\alpha_D(P) = \begin{cases} \alpha(P), P \notin T \\ 0, P \in T \end{cases} \qquad \alpha_E(P) = \begin{cases} 0, P \notin T \\ \alpha(P), P \in T \end{cases}.$$

It is obvious that $\alpha_D \in \mathcal{A}(D), \alpha_E \in \mathcal{A}(E)$ and hence $\alpha \in \mathcal{A}(D) + \mathcal{A}(E)$, which completes the first step as discussed above.

For the second step, recall that we have showed in the proof of Proposition 3.17 that for E of high enough degree one has $\mathcal{A}(E) + \mathbb{F}_q(C) = \mathcal{A}$. Since ω is nonzero, no such E is an element of S. Hence, the degree of divisors in S is bounded, and the proof is complete. □

Proposition 3.19 shows that one can associate to every differential ω a well-defined divisor K which we will from now on call the *divisor of* ω and denote by $\mathrm{div}(\omega)$. In analogy with the notion of principal divisors, any divisor that occurs as a divisor of a differential will be called a *canonical divisor*. Note that this also allows us to redefine $\Omega(D)$ in a way which is completely analogous to $\mathcal{L}(D)$, by putting $\Omega(D) = \{\omega \in \Omega : \mathrm{div}(\omega) \geq D\} \cup \{0\}$.

With the definition of canonical divisors, we can now state the main theorem of this section, the Serre duality theorem.

Theorem 3.20. Let D be a divisor of $C/\mathbb{F}_q$ and let K be a canonical divisor of $C/\mathbb{F}_q$. Then we have $\mathcal{L}(K-D) \cong \Omega(D)$ as $\mathbb{F}_q$-vector spaces.

We postpone the proof to the end of this section and begin by making a seemingly paradoxical observation about the statement of this theorem. Note that the $\Omega(D)$ does not depend on K, implying that the dimension of $\mathcal{L}(K - D)$ is independent of K. This is maybe surprising, but one should recall that if $D + \mathrm{div}(\varphi) = E$ then $\mathcal{L}(D) \cong \mathcal{L}(E)$, where the isomorphism is given by $\psi \mapsto \varphi\psi$. Hence, if we managed to prove that any two canonical divisors differ by a principal divisor, the above seeming paradox would be less surprising. Restating this in a different way, to show that differences of canonical divisors are principal divisors intuitively corresponds to showing that the "ratio" of the corresponding differentials is a rational function. Of course, we do not have a definition for the ratio of two linear maps $\omega_1, \omega_2 : \mathcal{A} \to \mathbb{F}_q$, but this restatement is very reminiscent of another idea in the theory of the Riemann surfaces.

Namely, one way to construct a meromorphic function on a Riemann surface is precisely by obtaining it as a ratio of two meromorphic differentials. In other words, one can directly show that the ratio of any two differentials on a Riemann surface is a meromorphic function. As always, it is good to recall that the connection to Riemann surfaces nothing more than an analogy, but still the intuition gained from recalling this approach can be very valuable.

Let us now show how to formalize the above discussion, showing that any two canonical divisors differ by a principal divisor. The main idea is to consider Ω as a $\mathbb{F}_q(C)$-vector space. Namely, if $\varphi \in \mathbb{F}_q(C)$ and $\omega \in \Omega$ we can define the differential $\varphi\omega$ by setting $\varphi\omega(\alpha) = \omega(\varphi\alpha)$ for any $\alpha \in \mathcal{A}$. It is not hard to check that with

[1] What is implicitly understood here is that we are considering the set of divisors as a partially ordered set under the relation $\leq$ and that the maximum of two divisors is simply their join. More explicitly, $\max\{D, E\}$ is a divisor F which satisfies $\nu_P(F) = \max\{\nu_P(D), \nu_P(E)\}$ for all $P \in \mathrm{PDiv}(C/\mathbb{F}_q)$.

this definition, the space Ω indeed becomes a $\mathbb{F}_q(C)$ vector space. What is more interesting, however, is that this vector space now has dimension 1 over $\mathbb{F}_q(C)$. In other words, for any two nonzero differentials ω_1, ω_2, there exists a rational function φ such that $\omega_1 = \varphi\omega_2$. Before we prove this, let us comment on how multiplying differentials by a rational function affects their divisors.

For any two nonzero rational functions $\varphi, \psi \in \mathbb{F}_q(C)$ we have $\operatorname{div}(\varphi\psi) = \operatorname{div}(\varphi) + \operatorname{div}(\psi)$, and therefore one would expect to have a similar statement when $\varphi \in \mathbb{F}_q(C)$ and $\omega \in \Omega$. And indeed, the following Proposition shows the direct analogy.

Proposition 3.21. For any nonzero rational function $\varphi \in \mathbb{F}_q(C)$ and any nonzero differential $\omega \in \Omega$ we have

$$\operatorname{div}(\varphi\omega) = \operatorname{div}(\varphi) + \operatorname{div}(\omega).$$

Proof. The first step will be to show that $\varphi\omega$ vanishes on $\operatorname{div}(\omega) + \operatorname{div}(\varphi)$. Let $D = \operatorname{div}(\omega)$ and $\alpha \in \mathcal{A}(D + \operatorname{div}(\varphi)) + \mathbb{F}_q(C)$. Since $\operatorname{div}(\alpha) + \operatorname{div}(\varphi) + D \geq 0$, we also have $\operatorname{div}(\varphi\alpha) + D \geq 0$, meaning $\omega(\varphi\alpha) = 0$. Hence, $\varphi\omega(\alpha) = 0$ and we conclude $\operatorname{div}(\varphi\omega) \geq \operatorname{div}(\omega) + \operatorname{div}(\varphi)$. By applying this inequality to φ^{-1} and $\varphi\omega$ we have $\operatorname{div}(\varphi^{-1}(\varphi\omega)) \geq \operatorname{div}(\varphi\omega) + \operatorname{div}(\varphi^{-1})$. Combining these two inequalities now yields the result. □

Proposition 3.22. The dimension Ω as a $\mathbb{F}_q(C)$-vector space is 1.

Proof. We need to show that for any two nonzero differentials ω_1, ω_2, there exists a rational function φ for which $\omega_1 = \varphi\omega_2$. In fact, we will show that there are two rational functions, φ_1, φ_2 such that $\varphi_1\omega_1 = \varphi_2\omega_2$, which definitely suffices.

To show this, we pick a divisor D such that both ω_1 and ω_2 vanish on $\mathcal{A}(D) + \mathbb{F}_q(C)$ (we know such divisors exist for both ω_1 and ω_2 separately, and we may take their minimum). Furthermore, we pick a large positive divisor G and construct two maps $T_1, T_2 : \mathcal{L}(D+G) \to \Omega(-G)$ given by $T_1 : \varphi \mapsto \varphi\omega_1$ and $T_2 : \varphi \mapsto \varphi\omega_2$. It is clear that $\varphi\omega_i$ is in $\Omega(-G)$, since $\operatorname{div}(\varphi\omega_i) = \operatorname{div}(\varphi) + \operatorname{div}(\omega_i) \geq -(D+G) + D = -G$ for $i = 1, 2$. Now, if we show that $\operatorname{im} T_1 \cap \operatorname{im} T_2 \neq \{0\}$, this automatically produces two nonzero functions φ_1, φ_2 with $\varphi_1\omega_1 = \varphi_2\omega_2$, completing the proof.

Checking that the images of T_1, T_2 intersect nontrivialy in now a simple exercise in dimension counting. Since both T_1, T_2 are injective, we have $\dim_{\mathbb{F}_q} \operatorname{im} T_1 = \dim_{\mathbb{F}_q} \mathcal{L}(D+G) \geq \deg D + \deg G - g$ by Riemann's inequality. Analogously we have $\dim_{\mathbb{F}_q} \operatorname{im} T_2 \geq \deg D + \deg G - g$. Finally, the $\mathbb{F}_q$-dimension of $\Omega(-G)$ can be computed from Proposition 3.17 giving:

$$\dim_{\mathbb{F}_q} \Omega(-G) = \dim_{\mathbb{F}_q} \mathcal{L}(-G) - \deg(-G) - 1 + g = \deg G + g - 1,$$

where we have used Corollary 3.8 to claim $\dim_{\mathbb{F}_q} \mathcal{L}(-G) = 0$. Hence, for big enough $\deg G$ we obtain

$$\dim_{\mathbb{F}_q} \operatorname{im} T_1 + \dim_{\mathbb{F}_q} \operatorname{im} T_2 - \dim_{\mathbb{F}_q} \Omega(-G) \geq \deg G + 2\deg D - 3g + 1 \geq 1,$$

which, as explained above, completes the proof. □

With Propositions 3.22 and 3.21, we can easily prove that the difference of any two canonical divisors is a principal divisor, a statement that we were interested in earlier. Namely, if ω_1, ω_2 are nonzero differentials, we have $\omega_1 = \varphi\omega_2$ for some $\varphi \in \mathbb{F}_q(C)$ and $\operatorname{div}(\omega_1) = \operatorname{div}(\varphi\omega_2) = \operatorname{div}(\varphi) + \operatorname{div}(\omega_2)$. Hence $\operatorname{div}(\omega_1) - \operatorname{div}(\omega_2) = \operatorname{div}(\varphi)$ is indeed principal.

With all of the above in mind, we are finally ready to prove Serre duality theorem, which is the last step towards the Riemann-Roch theorem.

Proof of the Theorem 3.20. Since K is a canonical divisor, one can pick a nonzero differential ω such that $\operatorname{div}(\omega) = K$. Now, consider the linear map between $T : \mathcal{L}(K - D) \to \Omega(D)$ given by $\varphi \mapsto \varphi\omega$. Note the equivalence $\operatorname{div}(\varphi\omega) \geq D \iff \operatorname{div}(\varphi) + (K - D) \geq 0$, which shows that $\varphi\omega$ is indeed in $\Omega(D)$.

To show that T is an isomorphism, we check that it is injective and surjective. The injectivity is clear - if $\varphi \in \ker T$ and $\varphi \neq 0$, we have $\omega = \varphi^{-1}\varphi\omega = \varphi^{-1}0 = 0$, a contradiction to the assumption that ω is nonzero. On the other hand, if $\omega' \in \Omega(D)$ is any differential, we have $\omega' = \varphi\omega$ for some $\varphi \in \mathbb{F}_q(C)$ by Proposition 3.22. Since $\varphi\omega \in \Omega(D)$, the equivalence presented above shows $\varphi \in \mathcal{L}(K - D)$, completing the proof that T is surjective.

Hence, T is an isomorphism and we have $\Omega(D) \cong \mathcal{L}(K - D)$. □

Proof of the Theorem 3.2. The Riemann-Roch theorem follows immediately as a combination of Proposition 3.17 and Theorem 3.20. □

Corollary 3.23. Let D be a divisor with $\deg D > 2g-2$. Then we have $\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g$.

Proof. From the Riemann-Roch theorem we have $\dim_{\mathbb{F}_q} \mathcal{L}(D) = \deg D + 1 - g + \dim_{\mathbb{F}_q} \mathcal{L}(K-D)$, where $K-D$ is a divisor of degree $\deg K - \deg D < 0$. Since the divisor of any rational function φ has $\deg \operatorname{div}(\varphi) = 0$, we cannot have $\operatorname{div}(\varphi) + K - D \geq 0$. Therefore, $\dim_{\mathbb{F}_q} \mathcal{L}(K-D) = 0$ and the statement of our corollary follows. □

3.6 History

The first steps towards the Riemann-Roch theorem were initiated by Riemann in the 1850s, in the study of meromorphic functions on Riemann surfaces. Namely, Riemann was interested in the spaces of meromorphic functions with bounded order of poles at certain points, and derived a version of what we have called Riemann's inequality. Almost a decade later, Riemann's student Roch extended the theorem further, pinning down the exact error term between $\dim_{\mathbb{C}} \mathcal{L}(D)$ and $\deg D + 1 - g$.

Since then, the Riemann Roch theorem has been extended to various other settings, such as the one presented in this thesis. The setting of the Riemann-Roch theorem in the function fields was first studies by F.K. Schmidt, with the goal of showing the rationality and functional equation, as we will do in the next chapter. Of course, the proof had to be significantly reworked, since the original approach relied heavily on the complex analytic structure of Riemann surfaces. Even though we have phrased the proof for curves over finite fields $\mathbb{F}_q$, the same argument shows the analogue of the Riemann-Roch theorem for any other perfect field.

Finally, let us remark that many analogues of the Riemann-Roch theorem were developed over the years, including an analogue related to graph theory and chip firing games introduces by Baker and Norine [7].

Chapter 4

Rationality and functional equation

The goal of this chapter is to show the first two assertions of Theorem 1.6, namely the rationality and the functional equation for the zeta function. As we will see, these two properties of zeta functions are much easier to show than the Riemann Hypothesis for curves, as they follow almost directly from the Riemann-Roch theorem. The high-level description of the argument is quite simple - first, we rewrite the zeta function of a curve in terms of the divisors of that curve, and then we apply the Riemann-Roch theorem to count these divisors. As before, we will fix a curve $C/\mathbb{F}_q$ and work with it throughout the whole chapter.

4.1 Picard group of a curve

In order to motivate the introduction of the Picard group, let us recall an analogous notion from algebraic number theory, the class group of a number field. Namely, if $K/\mathbb{Q}$ is a number field, with the ring of integers $\mathcal{O}_K$, we know that not every element of $\mathcal{O}_K$ may have a unique factorization into primes. The primary issue behind this is that there may be ideals of $\mathcal{O}_K$ which are not principal. However, one of the great results of classical algebraic number theory guarantees that every ideal can be uniquely factored into prime ideals.

Furthermore, one can define the equivalence classes of ideals by setting two ideals to be equivalent if one can be obtained from the other through a multiplication with a principal ideal. The set of equivalence classes can then be turned into a group under the standard ideal multiplication and the resulting group is called the class group $Cl(K)$. The class group plays the central role in describing many arithmetic properties of the number field K. An interested reader can find more information about the class group and their relation to number fields in [18].

To define a Picard group of a curve, we follow the same recipe applied to divisors rather than ideals. Recall that $\mathrm{Div}(C/\mathbb{F}_q)$ is the group of divisors of the curve $C/\mathbb{F}_q$ under the addition operation and $\mathrm{Princ}(C/\mathbb{F}_q)$ the subgroup of its principal divisors. Recall that we have also introduced a notion of equivalence between divisors in Definition 2.28 and that the equivalence classes of divisors in $\mathrm{Div}(C/\mathbb{F}_q)$ correspond exactly to elements of the quotient $\mathrm{Div}(C/\mathbb{F}_q)/\mathrm{Princ}(C/\mathbb{F}_q)$. Unlike the case of number fields, this quotient will not be finite for a very simple reason. Since all principal divisors have degree 0 and there exist divisors $D \in \mathrm{Div}(C/\mathbb{F}_q)$ of arbitrarily high degree, we cannot have only finitely many equivalence classes of divisors.

However, there is a simple way to remedy the situation. Consider the group of divisors of degree 0, denoted by $\mathrm{Div}^0(C/\mathbb{F}_q)$, and consider its quotient with $\mathrm{Princ}(C/\mathbb{F}_q)$. This object, the analogue of the class group of a number field, will indeed be finite, and we will use it to help us count the number of positive divisors of a given degree on $C/\mathbb{F}_q$.

Definition 4.1. The *Picard group* of the curve $C/\mathbb{F}_q$ is the defined as $\mathrm{Pic}(C/\mathbb{F}_q) = \mathrm{Div}^0(C/\mathbb{F}_q)/\mathrm{Princ}(C/\mathbb{F}_q)$.

To set up some notation, we denote the equivalence class of the divisor D by $[D]$.

Proposition 4.2. The Picard group of a curve is finite and we denote its order by h.

Proof. Our goal is to show that there are only finitely many equivalence classes of divisors of some degree $d > 0$,

say $\{[D_1], \ldots, [D_k]\}$ for some $k \geq 1$. This suffices because these classes are in one-to-one correspondence with the equivalence classes of degree 0, where the correspondence is given by subtracting $[D_1]$ from a given class.

If the degree d is at least g, say $d = g+1$, we can show that every divisor D of degree d is in the equivalence class of a positive divisor of degree d. Let D be a divisor of degree $d \geq g$, where we have $\dim_{\mathbb{F}_q} \mathcal{L}(D) \geq \deg D + 1 - g \geq 1$ by Riemann's inequality. Hence, there exists a rational function $\varphi \in \mathbb{F}_q(C)$ for which $D + \operatorname{div}(\varphi) \geq 0$. Then, the divisor $D + \operatorname{div}(\varphi)$ is positive and equivalent to D.

The last step of the proof is to show that there are only finitely many nonnegative divisors of degree d. But again, this is not hard to show. Since nonnegative divisors are sums of prime divisors, which all have degree at least 1, it suffices to check that there are only finitely many prime divisors of every given degree. But prime divisors of degree d are constituted of points of $\mathbb{P}^2(\mathbb{F}_{q^d})$ by Proposition 1.17. Hence, there are at most $q^{2d} + q^d + 1$ prime divisors of degree d, completing the proof. □

4.2 Divisors of degree 1

In this section, we will show that for any $d \in \mathbb{Z}$, there exists a divisor of degree d on the curve $C/\mathbb{F}_q$. Of course, this is equivalent to showing that there exists a divisor D of degree 1, since scaling D gives divisors of arbitrary degrees. Interestingly, the proof of this fact will be indirect and will rely on comparing the orders of poles at $t = 1$ of zeta functions associated to curves $C/\mathbb{F}_q$ and $C/\mathbb{F}_{q^m}$. We begin this section by showing how these zeta functions relate to each other.

Lemma 4.3. If $Z(C/\mathbb{F}_{q^m}, t)$ is the zeta function of the curve $C/\mathbb{F}_{q^m}$ and $Z(C/\mathbb{F}_q, t)$ is the zeta function of $C/\mathbb{F}_q$, then

$$Z(C/\mathbb{F}_{q^m}, t^m) = \prod_{k=1}^{m} Z(C/\mathbb{F}_q, e^{\frac{2i\pi k}{m}} t).$$

Proof. The key insight we will use in this proof is the relation between prime divisors of the curve $C/\mathbb{F}_{q^m}$ and the prime divisors of $C/\mathbb{F}_q$, as described in Proposition 1.17. More precisely, the proof of Proposition 1.19 shows that we have the following expression for $Z(C/\mathbb{F}_{q^m}, t^m)$:

$$Z(C/\mathbb{F}_{q^m}, t^m) = \prod_{P \in \mathrm{PDiv}(C/\mathbb{F}_{q^m})} \frac{1}{1 - t^{m \deg P}} = \prod_{Q \in \mathrm{PDiv}(C/\mathbb{F}_q)} \prod_{P \in \mathrm{PDiv}(C/\mathbb{F}_{q^m}), P \subset Q} \frac{1}{1 - t^{m \deg P}}.$$

The inner product can be evaluated easily, since Proposition 1.17 ensures that $\deg P = \frac{\deg Q}{(\deg Q, m)}$ and there are exactly $(\deg Q, m)$ divisors $P \subseteq Q$. Hence, we have

$$Z(C/\mathbb{F}_{q^m}, t^m) = \prod_{Q \in \mathrm{PDiv}(C/\mathbb{F}_q)} \left(1 - (t^{\deg Q})^{\frac{m}{(m, \deg Q)}}\right)^{-(\deg Q, m)} = \prod_{Q \in \mathrm{PDiv}(C/\mathbb{F}_q)} \prod_{\zeta = e^{2i\pi k/m}} \frac{1}{1 - (\zeta t)^{\deg Q}},$$

where the second equality comes from the basic properties of cyclotomic polynomials. This suffices to complete the proof because

$$Z(C/\mathbb{F}_{q^m}, t^m) = \prod_{\zeta = e^{2i\pi k/m}} \prod_{Q \in \mathrm{PDiv}(C/\mathbb{F}_q)} \frac{1}{1 - (\zeta t)^{\deg Q}} = \prod_{k=1}^{m} Z(C/\mathbb{F}_q, e^{\frac{2i\pi k}{m}} t).$$

□

As indicated previously, the main idea behind showing the existence of degree 1 divisor on $C/\mathbb{F}_q$ is to analyze the order of the pole of $Z(C/\mathbb{F}_q, t)$ at $t = 1$. However, in order to be able to do this, we need a rudimentary form of rationality of $Z(C/\mathbb{F}_q, t)$ which will allow us to analyze this pole. Hence, before dealing with degree 1 divisors, we present a general lemma about counting positive divisors, which will be used to establish this rudimentary form of rationality. In this lemma, and further throughout this chapter, the number of positive divisors of degree d on $C/\mathbb{F}_q$ is denoted by D_d.

Lemma 4.4. Let d be an integer and suppose that there exists a divisor of degree d on the curve $C/\mathbb{F}_q$. Then, we have the following relation between D_d and D_{2g-2-d}:

$$D_d = D_{2g-2-d} q^{d+1-g} + h \frac{q^{d+1-g} - 1}{q - 1}.$$

Proof. We will start the proof from the following simple observation. If D is a divisor of degree d, there exist exactly $\frac{q^{\dim \mathcal{L}(D)}-1}{q-1}$ nonnegative divisors E which are equivalent to D. This is relatively easy to see - if $E = D + \operatorname{div}(\varphi)$ is a nonnegative divisor, then φ must be a nonzero element of $\mathcal{L}(D)$. Moreover, if φ and ψ are not scalar multiples of each other, we have $\operatorname{div}(\varphi) \neq \operatorname{div}(\psi)$ and hence the resulting divisors E are different. Hence, we conclude that the number of nonnegative divisors E equivalent to D is simply the number of nonzero rational functions in $\mathcal{L}(D)$ up to scalar multiplication, just as claimed above.

Hence, we may express D_d as $\sum_D \frac{q^{\dim \mathcal{L}(D)}-1}{q-1}$, where the sum runs over the representatives of equivalence classes among degree d divisors. The Riemann-Roch theorem implies that $\dim \mathcal{L}(D) = d+1-g+\dim \mathcal{L}(K-D)$, where the divisor $K-D$ has degree $2g-2-d$. Moreover, as D runs through the representatives of the classes of degree d, the divisor $K-D$ runs over all classes of degree $2g-2-d$. Hence, the following simple computation completes the proof.

$$D_k = \sum_D \frac{q^{\dim \mathcal{L}(D)}-1}{q-1} = \sum_D \frac{q^{d+1-g+\dim \mathcal{L}(K-D)}-1}{q-1} =$$
$$= \sum_{E=K-D} \frac{q^{d+1-g}(q^{\dim \mathcal{L}(E)}-1)+q^{d+1-g}-1}{q-1} = D_{2g-2-d}q^{d+1-g} + h\frac{q^{d+1-g}-1}{q-1}.$$

□

Corollary 4.5. If $d > 2g-2$ and there exists a divisor of degree d on $C/\mathbb{F}_q$, the number of nonnegative divisors of degree d is $D_d = h \cdot \frac{q^{d-g+1}-1}{q-1}$, where h is the cardinality of the Picard group of C.

Proof. Applying Lemma 4.4 and noting that there are no nonnegative divisors of degree $2g-2-d<0$ suffices to prove this corollary. □

Finally, the stage is set and we are ready to prove the main result of this section - the existence of degree 1 divisors. The proof of this result, which only talks about the divisors on a curve, uses the properties of zeta functions to deduce this. This is an interesting interplay between the analytic properties of zeta functions and algebraic properties of the curves over a finite field.

Proposition 4.6. There exists a divisor of degree 1 on the curve $C/\mathbb{F}_q$.

Proof. Consider the degree map, which takes divisors to integers, $\deg : \operatorname{Div}(C/\mathbb{F}_q) \to \mathbb{Z}$ and consider its image in $\mathbb{Z}$. Clearly, the image is a subgroup of $\mathbb{Z}$, generated by an number $m > 0$. Our goal is to show that $m = 1$. The way we do this is by considering the pole of the zeta function of $C/\mathbb{F}_q$ and $C/\mathbb{F}_{q^m}$ at $t = 1$ and showing that both of these must be simple poles. We will then show that the zeta function of $C/\mathbb{F}_{q^m}$ depends only on t^m, and obtain a contradiction using Lemma 4.3. Let us now present the details of the argument.

Our first step will be to obtain a formula for $Z(C/\mathbb{F}_q, t)$, using the expression (1.6). For $d = km$ and $d > 2g-2$ by Corollary 4.5 we have $D_d = h\frac{q^{d-g+1}-1}{q-1}$ and therefore

$$Z(C/\mathbb{F}_q, t) = \sum_{d=0, m|d}^{2g-2} D_d t^d + \sum_{d>2g-2, m|d} h\frac{q^{d-g+1}-1}{q-1}t^d = F(t^m) + \frac{h}{q-1}\left(\frac{q^{d_0}t^{d_0}}{1-q^m t^m} - \frac{t^{d_0}}{1-t^m}\right),$$

where d_0 is the smallest multiple of m bigger than $2g-2$ and F is some polynomial. Therefore, $Z(C/\mathbb{F}_q, t)$ can be associated with a rational functions with a simple pole at $t = 1$. Analogously, we have that $Z(C/\mathbb{F}_{q^m}, t)$ is a rational function with a simple pole at $t = 1$, and therefore $Z(C/\mathbb{F}_{q^m}, t^m)$ also has a simple pole at $t = 1$. However, Lemma 4.3 states

$$Z(C/\mathbb{F}_{q^m}, t^m) = \prod_{k=1}^{n} Z(C/\mathbb{F}_q, e^{2i\pi k/m}t).$$

The main observation is that $Z(C/\mathbb{F}_q, t)$ depends only on t^m and therefore we have $Z(C/\mathbb{F}_q, e^{2i\pi k/m}t) = Z(C/\mathbb{F}_q, t)$ for all $k \in \{1, \ldots, n\}$. We conclude that $Z(C/\mathbb{F}_{q^m}, t^m) = Z(C/\mathbb{F}_q, t)^m$. However, both functions have a simple pole at $t = 1$, and therefore $m = 1$. This completes the proof. □

4.3 Rationality of the zeta function

In this section, we establish rationality of $Z(C/\mathbb{F}_q, t)$ in the form given by equation (1.1).

Proposition 4.7. The zeta function is a rational function and it can be written in the form $Z(C/\mathbb{F}_q, t) = \frac{L(t)}{(1-t)(1-qt)}$, where $L(t)$ is a polynomial of degree at most $2g$ with integer coefficients.

Proof. The main idea behind the proof is to exploit the divisor counting formula from Corollary 4.5 along with the alternative expression for $Z(C/\mathbb{F}_q, t)$ presented in Proposition 1.19 to compute $Z(C/\mathbb{F}_q, t)$. More precisely, we have

$$Z(C/\mathbb{F}_q, t) = \sum_{d\geq 0} D_d t^d = \sum_{d=0}^{2g-2} D_d t^d + \sum_{d\geq 2g-1} h\frac{q^{d-g+1}-1}{q-1} t^d.$$

We will denote the first sum by $F(t)$, for a general polynomial F of degree at most $2g-2$, and we will explicitly compute the remaining sum:

$$\sum_{d\geq 2g-1} \frac{q^{d-g+1}-1}{q-1} t^d = \frac{t^{2g-1}}{q-1}\sum_{k\geq 0}\left(q^{g+k}t^k - t^k\right) = \frac{t^{2g-1}}{q-1}\left(\frac{q^g}{1-qt} - \frac{1}{1-t}\right) = \frac{t^{2g}(q-q^g) + t^{2g-1}(q^g-1)}{(q-1)(1-t)(1-qt)}.$$

Hence, we obtain the following expression for $Z(C/\mathbb{F}_q, t)$, which evidently verifies the statement of this proposition.

$$Z(C/\mathbb{F}_q, t) = \frac{ht^{2g}\frac{q-q^g}{q-1} + ht^{2g-1}\frac{q^g-1}{q-1} + F(t)(1-t)(1-qt)}{(1-t)(1-qt)}. \tag{4.1}$$

□

Remark 4.8. The proof given above given allows us to be even more precise about the polynomial $L(t)$. Namely, using the Riemann-Roch theorem, we can easily calculate that there are $(h-1)\frac{q^{g-1}-1}{q-1} + \frac{q^g-1}{q-1}$ nonnegative divisors of degree $2g-2$, since $\dim_{\mathbb{F}_q}\mathcal{L}(D) = g-1$ whenever D is not a canonical divisor and $\dim_{\mathbb{F}_q}\mathcal{L}(D) = g$ for canonical divisors. In turn, this expression corresponds to the coefficient of t^{2g-2} in $F(t)$. Hence, computing the coefficient of t^{2g} in $L(t)$ from the expression (4.1) gives that the leading coefficient of $L(t)$ is q^g.

4.4 Functional equation

The goal of this section is to prove the functional equation (1.2) for $Z(C/\mathbb{F}_q, t)$.

Proposition 4.9. The zeta function $Z(C/\mathbb{F}_q, t)$ satisfies the functional equation $Z(C/\mathbb{F}_q, \frac{1}{qt}) = q^{1-g}t^{2-2g}Z(C/\mathbb{F}_q, t)$.

Proof. The proof proceeds in three major steps. First, we will reformulate the functional equation for $Z(C/\mathbb{F}_q, t)$ into an equivalent functional equation for $L(t)$, and find a condition on the coefficients of the polynomial $L(t)$ equivalent to it. Secondly, we will obtain a formula for the coefficients of $L(t)$ in terms of the nonnegative divisors of the curve $C/\mathbb{F}_q$. Finally, we will then use the appropriate symmetry of divisor counts coming from Lemma 4.4 to verify this condition on the coefficients of $L(t)$. Let us now present the details of the proof.

We begin by reformulating the functional equation for $Z(C/\mathbb{F}_q, t)$ into a condition for the coefficients of the polynomial $L(t)$. Recall that the zeta function is rational and that it has the form $Z(C/\mathbb{F}_q, t) = \frac{L(t)}{(1-t)(1-qt)}$. The denominator of this fraction $(1-t)(1-qt)$ transforms very predictably under the map $t \mapsto 1/qt$, and we have $(1-\frac{1}{t})(1-\frac{1}{qt}) = t^{-2}q^{-1}(1-t)(1-qt)$. Hence, the functional equation for $Z(C/\mathbb{F}_q, t)$ is equivalent to the following functional equation for $L(t)$:

$$L\left(\frac{1}{qt}\right) = q^{-g}t^{-2g}L(t). \tag{4.2}$$

Writing the polynomial $L(t)$ as $L(t) = \sum_{d=0}^{2g} a_d t^d$, we find that the above equation is equivalent to

$$\sum_{d=0}^{2g} a_d q^{-d} t^{-d} = \sum_{d=0}^{2g} a_d t^{d-2g} q^{-g}.$$

By comparing the corresponding coefficients, we see this polynomial identity is equivalent to having $a_d q^{-d} = a_{2g-d}q^{-g}$, i.e. $a_d = a_{2g-d}q^{d-g}$ for all $d \in \{0, \ldots, 2g\}$. Hence, showing the functional equation for $L(t)$ is equivalent to verifying that $a_d = a_{2g-d}q^{d-g}$ for all $d \in \{0, \ldots, 2g\}$.

The coefficients a_d of the polynomial D_d are very closely connected to the number of nonnegative divisors of degree d, which we can see from the proof of Proposition 4.7. More precisely, we have

$$L(t) = ht^{2g}\frac{q-q^g}{q-1} + ht^{2g-1}\frac{q^g-1}{q-1} + F(t)(1-t)(1-qt),$$

where $F(t) = \sum_{d=0}^{2g-2} D_d t^d$. Hence, for $d \in \{2,\dots,2g-2\}$ we have $a_d = D_{d-2}q - D_{d-1}(q+1) + D_d$ and the remaining coefficients can be computes as $a_0 = D_0, a_1 = D_1 - (q+1)D_0$, $a_{2g} = q^g$ (recall the computation in Remark 4.8) and $a_{2g-1} = qD_{2g-3} - (q+1)D_{2g-2} + h\frac{q^g-1}{q-1}$.

Let us check the condition $a_d = a_{2g-d}q^{d-g}$ first for $d \in \{2,\dots,2g-2\}$. In terms of the divisor counts, we can express this equivalently as $qD_{d-2} - (q+1)D_{d-1} + D_d = q^{d-g}(D_{2g-2-d} - (q+1)D_{2g-1-d} + D_{2g-d})$. These relations follow almost immediately from Lemma 4.4 applied to $d-2, d-1$ and d through the following computation

$$\begin{aligned}
&qD_{d-2} - (q+1)D_{d-1} + D_d = \\
&= q\left(D_{2g-d}q^{d-1-g} + h\frac{q^{d-1-g}-1}{q-1}\right) - (q+1)\left(D_{2g-1-d}q^{d-g} + h\frac{q^{d-g}-1}{q-1}\right) + \left(D_{2g-d}q^{d+1-g} + h\frac{q^{d+1-g}-1}{q-1}\right) \\
&= q^{d-g}(D_{2g-d} + D_{2g-1-d} + qD_{2g-d}) + \frac{h}{q-1}\Big(q^{d-g} - q - (q+1)(q^{d-g}-1) + q^{d+1-g} - 1\Big),
\end{aligned}$$

where the second term in the final line is easily seen to be zero.

Now, we are left with only two things to check - that $a_0 = q^{-g}a_{2g}$ and that $a_1 = q^{1-g}a_{2g-1}$. The first one is trivial - we know that $a_0 = D_0 = 1$ (since 0 is the only nonnegative divisor of degree 0) and $a_{2g} = q^g$. Finally, to show $a_1 = q^{1-g}a_{2g-1}$ we have another application of Lemma 4.4

$$\begin{aligned}
a_1 = D_1 - (q+1)D_0 &= D_{2g-3}q^{2-g} - (q+1)D_{2g-2}q^{1-g} + \frac{h}{q-1}\left(\frac{q^{2-g}-1}{q-1} - (q+1)\frac{q^{1-g}-1}{q-1}\right) \\
&= q^{1-g}\left(qD_{2g-3} - (q+1)D_{2g-2} + h\frac{q - q^{g-1} - (q+1)(1-q^{g-1})}{q-1}\right) = q^{1-g}a_{2g-1}.
\end{aligned}$$

□

Chapter 5

Schmidt's approach to Stepanov's method

In this chapter, we will depart from all the machinery we developed in the previous chapters and present an elementary proof of the Riemann Hypothesis for curves. Our approach will be based on the polynomial method as presented by Schmidt [25].

Unlike in previous sections, we will work in affine space ${\mathbb{F}_q}^2$, where $q = p^k$ is a power of a prime, and consider an absolutely irreducible polynomial $f(X, Y) \in \mathbb{F}_q[X, Y]$ of degree d. We will also adopt a notational convention that capital letters, such as X, Y, denote the variables while the lowercase ones, such as x, y, refer to the values of these variables. Finally, since we will be working with polynomials in several variables, e.g. $F(X_1, \dots, X_k)$, we need a notation convention for their degrees with respect to a subset of variables. Therefore, we will be denoting the degree of F with respect to variables $X_1, \dots, X_\ell$ by $\deg_{X_1,\dots,X_\ell} F$.

We will show the following theorem.

Theorem 5.1. If $N = \#C/\mathbb{F}_q$ is the number of solutions to the equation $f(x, y) = 0$ with $x, y \in \mathbb{F}_q$, where $f \in \mathbb{F}_q[X, Y]$ is a polynomial of degree d, then

$$|N - q| \leq O\left(q^{1/2} d^3\right). \tag{5.1}$$

Combined with Proposition 1.12, this implies the Riemann hypothesis for curves, which itself gives a much better bound $|N - q| \leq 2gq$, where g is the genus of the corresponding curve. Hence, the exact constant and the power of d in the bound (5.1) is not crucial, and the most important thing is that the bound is proportional to $q^{1/2}$. Therefore, in order to improve the clarity of the argument, our bounds will be slightly cruder than those presented by Schmidt in [25], but this has no effect on the final conclusion as described above.

For the proof of this theorem, we use the polynomial method. The basic idea is that we will construct a nonzero polynomial of low degree which will vanish at all $\mathbb{F}_q$-rational points of $C/\mathbb{F}_q$. Since the number of points where the polynomial can vanish is bounded by its degree, this provides a bound on the number of $\mathbb{F}_q$-rational points of $C/\mathbb{F}_q$. To construct the nonzero auxiliary polynomial, we will consider its coefficients as unknowns and consider the vanishing at points of $C/\mathbb{F}_q$ as linear constraints on these coefficients. Then, a linear algebra argument will guarantee that if we have more unknowns than constraints, the homogeneous system of equations has a nonzero solution, which will construct our nonzero polynomial. In our particular case, the auxiliary polynomial constructed in this manner will be in two variables, and we will need to transform it slightly, eliminate one of the variables, and prove that the polynomial remains nonzero after this transformation.

Let us now describe the steps of the proof is a slightly more formal manner.

- Step 1: We will begin by preprocessing the polynomial f to show that without loss of generality we can assume $f(X, Y) = Y^d + g_1(X)Y^{d-1} + \dots + g_d(X)$ where $\deg g_i(X) \leq i$ and $\partial_Y f$ is not a zero polynomial.

- Step 2: For a fixed $x \in \mathbb{F}_q$, consider the roots of $f^{(x)}(Y) = f(x, Y)$ as a univariate polynomial in Y. If all of its roots are distinct, let $I_1(x)$ be the set of roots in $\mathbb{F}_q$ and $I_2(x)$ be the set of roots not in $\mathbb{F}_q$. We will construct polynomials $e_1(X, Y, Y')$ of degree $d_1 = 1$ and $e_2(X, Y, Y')$ of degree $d_2 = d - 1$ such that for all $y \in I_1(x)$ we have $e_1(x, y, y^q) = 0$ and for all $y \in I_2(x)$ we have $e_2(x, y, y^q) = 0$.

- Step 3: On the set of rational functions in two variables, $\mathbb{F}_q(X,Y)$, we define a map D by the following equation
$$Dg(X,Y) = \partial_X g(X,Y) - \frac{\partial_X f(X,Y)}{\partial_Y f(X,Y)} \partial_Y g(X,Y).$$[1]
We will use the property of $I_t(x)$ (where $t \in \{1,2\}$) established in Step 2 to construct a polynomial $c_t(X,Y)$, not divisible by $f(X,Y)$, which satisfies $\deg c_t(X,Y) \leq \frac{d_t}{d} qM + q(d-1)$ and $D^l c_t(x,y) = 0$ for all $x \in \mathbb{F}_q, y \in I_t(x), l \in \{0,\dots,M-1\}$ (here, M will be a parameter left to be chosen later). This step is where we employ our linear algebraic argument together with dimension counting to construct the auxiliary polynomial.

- Step 4: We will convert the polynomial $c_t(X,Y)$ into a nonzero univariate polynomial $h_t(x)$ of degree at most $d_t qM + qd(d-1)$ such that h_t has a zero of order $M|I_t(x)|$ at every point $x \in \mathbb{F}_q$. Then, by bounding the sum of vanishing multiplicities of a polynomial $h_t(X)$ we obtain the needed bound on $\sum_{x\in\mathbb{F}_q} |I_t(x)|$, i.e. on the total number of roots of $f(X,Y)$.

In the attempt to execute this program, we will run into several difficulties. We require the auxiliary polynomial h_t we construct to vanish to order M at various points of ${\mathbb{F}_q}^2$ and we set $M \sim q^{1/2}/d$. Suppose for a moment that $M < p$, where p is the characteristic of the underlying field. Then, we claim that the sum of vanishing multiplicities at points of the finite field is at most $\deg h_t$, which is proved by showing that if first M derivatives of h_t vanish at x, then $(X-x)^M | h_t$.

However, note that $q = p^k$ and therefore p could be smaller than M. Then, the natural question arises - what does it mean for a polynomial to vanish to order $M > p$ at a point? Our previous definition, that a polynomial vanishes at x_0 to order M is the first $M-1$ derivative vanish, is not sufficient anymore. The reason for this is that the derivatives of order bigger than p always vanish identically, and hence we cannot hope to show $(x-x_0)^M | h_t$ in this setting. Essentially, the problem is that the usual derivatives can accurately capture only the orders of zeros up to p. Hence, we need a new definition of the derivative, that will take into account the positive characteristic we are working with. We call this new notion a hyperderivative, and we introduce it in the Section 5.1.

5.1 Setup: Hyperderivatives

The main idea behind the definition of the hyperderivatives will be simple enough - we will define the ℓ-th hyperderivative of a polynomial $r \in \mathbb{F}_q[X]$ as $\frac{1}{\ell!}$ times the usual derivative, where the factor $\frac{1}{\ell!}$ is intended to cancel out factors which make the higher-order derivatives zero in positive characteristic. Of course, the difficulty is that if $\ell \geq p$, we are dividing by zero since $\ell! = 0$ in $\mathbb{F}_q$. To avoid dividing by zero, we will pretend we work in characteristic zero before we cancel out the corresponding powers of p, and then come back to $\mathbb{F}_q$. Let us now formulate more precisely the procedure we use.

The main ingredient that will allow us to calculate without worrying about positive characteristic will be lifting to the characteristic zero fields. Namely, recall that the p-adic valuation ν_p on $\mathbb{Q}$ introduced in the example 2.8. We will denote by $R_\mathbb{Q}$ the valuation ring of this valuation, and by $\mathfrak{M}_\mathbb{Q}$ its maximal ideal (i.e. the set of elements with valuation at least 1). The residue field of ν, $F_\mathbb{Q}$ is simply $\mathbb{F}_p$ and we have a natural projection map $\pi_\mathbb{Q} : R_\mathbb{Q} \to F_\mathbb{Q} = R_\mathbb{Q}/M_\mathbb{Q}$. In general, for a field K with a valuation ν_p on it, we will denote by R_K its valuation ring, $\mathfrak{M}_G$ its maximal ideal and by F_K the residue field, as introduced in section 2.2.

[1] At first sight, this definition of D may seem completely unmotivated and we will try to provide some motivation for it in this footnote. The main idea behind this definition is that we want to measure the order of vanishing of a polynomial g along a curve by considering its derivatives. Hence, it is natural to take derivatives of the polynomial g along the curve defined by $f(x,y) = 0$, instead of taking the partial derivatives in all directions. If we imagine that we are working over the complex numbers, say with $(x,y) \in \mathbb{C}^2$ and we have $\partial_Y f(x,y) \neq 0$, the implicit function theorem guarantees that y can be written as an analytic function of x, $y = y(x)$. Then, taking the derivative of f along the curve $f(x,y) = 0$ correspond to differentiating $g(x,y(x))$ with respect to x, which gives $Dg(x,y(x)) = \partial_X g(x,y(x)) + \partial_Y g(x,y(x))y'(x)$. Finally, $y'(x)$ can be computed from the equation $Df(x,y) = \partial_X f(x,y) + \partial_Y f(x,y)y'(x) = 0$, giving $y'(x) = -\frac{\partial_X f(x,y)}{\partial_Y f(x,y)}$. Plugging this back into the formula for $Dg(x,y)$ gives exactly the above formula for Dg. Of course, it is important to keep in mind that our definition works over finite fields, without any relation to analytic derivatives which only served as a inspiration.

Consider the number field $K = \mathbb{Q}(\widehat{\zeta})$ which is obtained by adjoining a primitive $(q-1)$-th root of unity $\widehat{\zeta}$ to $\mathbb{Q}$. If $[\mathbb{Q}(\widehat{\zeta}) : \mathbb{Q}] = \delta$, we may write every element of K as $a_0 + a_1\widehat{\zeta} + \cdots + a_{\delta-1}\widehat{\zeta}^{\delta-1}$, for some $a_0, \ldots, a_{\delta-1} \in \mathbb{Q}$. Our initial observation is that we can extend ν_p to the field K by setting

$$\nu_p(a_0 + a_1\widehat{\zeta} + \cdots + a_{\delta-1}\widehat{\zeta}^{\delta-1}) = \min\{\nu_p(a_0), \ldots, \nu_p(a_{\delta-1})\}. \tag{5.2}$$

As we will later show, this formula defines a proper valuation on K, whose residue field is $\mathbb{F}_q$.

The next step will be to extend the valuation ν_p to the field of rational functions $L = K(X, Y_1, \ldots, Y_d)$ in a similar way, and show that the corresponding residue field is $F_L = \mathbb{F}_q(X, Y_1, \ldots, Y_d)$. We will show this by induction on the number of variables. This will also induce projection maps $R_L \mapsto \mathbb{F}_q(X, Y_1, \ldots, Y_d)$, and pulling back along these maps will allow us to compute the derivatives in characteristic zero. Schematically, we will have the following diagram.

$$\begin{array}{ccccc}
\mathbb{Q} & \hookrightarrow & K = \mathbb{Q}(\widehat{\zeta}) & \hookrightarrow & L = K(X, Y_1, \ldots, Y_d) \\
\uparrow & & \uparrow & & \uparrow \\
R_{\mathbb{Q}} & \hookrightarrow & R_K & \hookrightarrow & R_L \\
\downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\
F_{\mathbb{Q}} = \mathbb{F}_p & \hookrightarrow & F_K = \mathbb{F}_q & \hookrightarrow & F_L = \mathbb{F}_q(X, Y_1, \ldots, Y_d)
\end{array}$$

Before we state the results of this section, let us establish some conventions. The image φ under the map $\pi : \varphi \mapsto \widehat{\varphi}$ of an element $\widehat{\varphi}$ in valuation ring is called its *projection*, while $\widehat{\varphi}$ is called a *lift* of φ. Furthermore, the variables with hats will refer to the lifts, while the variables without hats will refer to elements of the residue field. Let us now check that the valuations in question do extend as we expect them to. First, we deal with the case of adjoining a variable.

Proposition 5.2. Let G be a field with a valuation ν_p. If $G(X)$ is the field of rational functions in a single variable X with coefficients in G, the valuation ν_p can be extended to $G(X)$ by setting

$$\nu_p\left(\frac{a_0 + a_1X + \cdots + a_nX^n}{b_0 + b_1X + \cdots + b_mX^m}\right) = \min\{\nu_p(a_0), \ldots, \nu_p(a_n)\} - \min\{\nu_p(b_0), \ldots, \nu_p(b_m)\}.$$

Moreover, the residue field of this valuation is $F_{G(X)} = F_G(X)$ and the projection map $\pi : R_{G(X)} \to F_{G(X)}$ extends the existing map $\pi : R_G \to F_G$.

Proof. The essence of this proof is just carefully checking that all properties of ν_p are maintained under this new definition. Throughout the proof, a general rational function will be denoted by $\varphi = \frac{g}{h}$, where $g(x) = a_0 + a_1X + \cdots + a_nX^n$, $h(x) = b_0 + b_1X + \cdots + b_mX^m$.

It is obvious that $\nu_p : G(X) \to \mathbb{Z} \cup \{\infty\}$ is surjective, since it is already surjective on G. Furthermore, if $\nu_p(\varphi) = 0$, we must have $\min\{\nu_p(a_i)\} = \infty$, meaning that $g(X) \equiv 0$ and so $\varphi = 0$. Finally, note that it suffices to check the properties $\nu_p(\varphi_1\varphi_2) = \nu_p(\varphi_1) + \nu_p(\varphi_2)$ and $\nu_p(\varphi_1 + \varphi_2) \geq \min\{\nu_p(\varphi_1), \nu_p(\varphi_2)\}$ just on polynomials, since the valuation extends additively to all rational functions.

Let us now take two polynomials $g, h \in G[X]$, and an element $t \in G$ with $\nu_p(t) = 1$.[2] To check $\nu_p(g + h) \geq \min\{\nu_p(g), \nu_p(h)\}$ is easy, since

$$\begin{aligned}
\nu_p(g + h) &= \min\{\nu_p(a_0 + b_0), \ldots, \nu_p(a_n + b_n)\} \\
&\geq \min\{\min\{\nu_p(a_0), \nu_p(b_0)\}, \ldots, \min\{\nu_p(a_n), \nu_p(b_n)\}\} = \min\{\nu_p(g), \nu_p(h)\}.
\end{aligned}$$

We will now check that $\nu_p(gh) = \nu_p(g) + \nu_p(h)$. By multiplying g, h by an appropriate power of t, we may assume $\nu_p(g) = \nu_p(h) = 0$. Then, the goal is to show that $\nu_p(gh) = 0$. We have $a_0, \ldots, a_m, b_0, \ldots, b_n \in R_G$ and so all coefficients of gh are also in R_G, meaning $\nu_p(gh) \geq 0$. To show the reverse bound, choose minimal indices i, j for which $\nu_p(a_i) = 0, \nu_p(b_j) = 0$. Then, the coefficient of X^{i+j} in gh has valuation 0, by the strict triangle inequality, and hence $\nu_p(gh) = 0$.

[2] In our case, t can simply be the rational prime p, but since the Proposition is phrased for general fields, we prefer to use notation in line with Section 2.2.

Finally, the projection map to the residue field is defined by $\pi(a_0 + \cdots + a_n X^n) = \pi(a_0) + \cdots + \pi(a_n)X^n$ on the polynomials, and extended multiplicatively to the rational functions. It is not hard to check that the arising residue field is then $F_G(X)$. □

Proposition 5.2 guarantees that we can extend a valuation ν to the field of rational functions. We can do the same with an algebraic extension of a field, almost by the same argument. Since we only need to apply this once, to extend ν_p from $\mathbb{Q}$ to $\mathbb{Q}(\widehat{\zeta})$, we will prove it only in this case, although the general case follows easily too.

Proposition 5.3. The valuation ν_p extends from $\mathbb{Q}$ to $K = \mathbb{Q}(\widehat{\zeta})$, where $\widehat{\zeta}$ is a primitive $(q-1)$-th root of unity, by setting $\nu_p(\widehat{\zeta}) = 0$. Moreover, the residue field of this valuation is $F_K = \mathbb{F}_q$ and the projection map $\pi : R_K \to \mathbb{F}_q$ extends the existing map $\pi : R_{\mathbb{Q}} \to \mathbb{F}_p$.

Proof. Similarly to the proof of Proposition 5.2, we extend the valuation ν_p by defining $\nu_p(a_0 + a_1\widehat{\zeta} + \cdots + a_{\delta-1}\widehat{\zeta}^{\delta-1}) = \min\{\nu_p(a_0), \ldots, \nu_p(a_{\delta-1})\}$, where $\delta = [K : \mathbb{Q}]$.

The main thing we need to check in this proof is that the residue field $F_K = \mathbb{F}_q$, since the checking that the valuation and the projection map extend is almost identical to the argument given in the proof of Proposition 5.2.

Say under the projection map $\pi : R_K \mapsto F_K$, the element $\widehat{\zeta}$ is mapped to ζ. Then, ζ is still a primitive $q-1$-th root of unity in F_K and F_K is generated by the projections of element of $\mathbb{Q}$ and ζ. Hence, $F_K = \mathbb{F}_p(\zeta) = \mathbb{F}_q$. □

Hence, by applying Proposition 5.3 first and then Proposition 5.2 repeatedly, we conclude that the valuation ν_p extends from $\mathbb{Q}$ to L.

The reason for the discussion of extensions of these valuations is for us to able to define hyperderivatives, which we are now ready to do. If $\widehat{f}(X, Y)$ is a fixed lift of $f(X, Y)$ of degree d and $\widehat{\varphi}$ an arbitrary rational function of $K(X, Y_1, \ldots, Y_d)$, we define a linear operator D by setting

$$D\widehat{\varphi}(X, Y_1, \ldots, Y_d) = \partial_X \widehat{\varphi}(X, Y_1, \ldots, Y_d) - \sum_{i=1}^{d} \frac{\partial_X \widehat{f}(X, Y_i)}{\partial_Y \widehat{f}(X, Y_i)} \partial_{Y_i} \widehat{\varphi}(X, Y_1, \ldots, Y_d).$$

Then, the ℓ-th hyperderivative of a rational function $\widehat{\varphi} \in L$ is defined as $E^{(\ell)}\widehat{\varphi} = \frac{1}{\ell!} D^\ell \widehat{\varphi}$.

The hyperderivatives $E^{(\ell)}$ are linear operators satisfying many of the familiar properties of derivatives. The following proposition shows that some well known formulas for derivatives even simplify in the case of hyperderivatives.

Proposition 5.4. Let $\widehat{r}_1, \ldots, \widehat{r}_k$ be arbitrary polynomials in $L = K[X, Y_1, \ldots, Y_d]$, and let $\widehat{s} = \prod_{i=1}^{k} \widehat{r}_i$. Then, we have

$$E^{(\ell)}\widehat{s} = \sum_{\ell_1 + \cdots + \ell_k = \ell} E^{(\ell_1)}\widehat{r}_1 \cdots E^{(\ell_k)}\widehat{r}_k.$$

Proof. Recalling that $E^{(\ell)} = \frac{1}{\ell!} D^\ell$, one can rewrite this equation as

$$D^\ell \widehat{s} = \sum_{\ell_1 + \cdots + \ell_k = \ell} \frac{\ell!}{\ell_1! \cdots \ell_k!} D^{\ell_1}\widehat{r}_1 \cdots D^{\ell_k}\widehat{r}_k.$$

Of course, this relation follows immediately from the product rule for derivatives, and hence it suffices to show that $D(\widehat{r}_1\widehat{r}_2) = \widehat{r}_1 D\widehat{r}_2 + \widehat{r}_2 D\widehat{r}_1$ for arbitrary polynomials $\widehat{r}_1, \widehat{r}_2$. Since the operator D is just a linear combination of partial derivatives, the product rule follows easily

$$\begin{aligned} D(\widehat{r}_1\widehat{r}_2) &= \partial_X(\widehat{r}_1\widehat{r}_2) - \sum_i \frac{\partial_X \widehat{f}(X, Y_i)}{\partial_Y \widehat{f}(X, Y_i)} \partial_{Y_i}(\widehat{r}_1\widehat{r}_2) \\ &= \widehat{r}_1 \partial_X \widehat{r}_2 + \widehat{r}_2 \partial_X \widehat{r}_1 - \sum_i \frac{\partial_X \widehat{f}(X, Y_i)}{\partial_Y \widehat{f}(X, Y_i)} (\widehat{r}_1 \partial_{Y_i} \widehat{r}_2 + \widehat{r}_2 \partial_{Y_i} \widehat{r}_1) \\ &= \widehat{r}_1 D\widehat{r}_2 + \widehat{r}_2 D\widehat{r}_1. \end{aligned}$$

□

Let us now show how to define hyperderivatives of functions in the residue field. If $\varphi \in \mathbb{F}_q(X, Y_1, \ldots, Y_d)$ is a rational function with a lift $\widehat{\varphi} \in L$ satisfying $\pi(\widehat{\varphi}) = \varphi$, the ℓ-th hyperderivative of φ is defined by

$$E^{(\ell)}(\varphi) = \pi(E^{(\ell)}\widehat{\varphi}). \tag{5.3}$$

A priori, it is not obvious that $E^{(\ell)}(\varphi)$ does not depend on the choice of the lift $\widehat{\varphi}$ or that $E^{(\ell)}\widehat{\varphi}$ is indeed in the valuation ring. Hence, we present the proof of this fact in the form of the following proposition.

Proposition 5.5. Equation (5.3) gives a well-defined linear operator $E^{(\ell)}$ on $\mathbb{F}_q(X, Y_1, \ldots, Y_d)$.

Proof. As previously indicated, we have to check two things - that hyperderivatives of a lift are still in the valuation ring (so that we may take their projection) and that $E^{(\ell)}\varphi$ does not depend on the lift $\widehat{\varphi}$.

Let us first show that for any choice of $\widehat{\varphi} \in R_L$, where $L = K(X, Y_1, \ldots, Y_d)$, we have $E^{(\ell)}\widehat{\varphi} \in R_L$, i.e. that $\nu_p(E^{(\ell)}\widehat{\varphi}) \geq 0$. We do this in steps. The first step will deal with the case when $\widehat{\varphi}$ is a polynomial in X. Then, we will check that the statement holds for general polynomials $\widehat{\varphi} \in R_L \cap K[X, Y_1, \ldots, Y_d]$, and only then we will extend the result to all rational functions.

Suppose $\widehat{\varphi} \in \mathbb{F}_q[X]$ with the form $\widehat{\varphi}(X) = \sum_{k=0}^{\deg \widehat{\varphi}} \widehat{\lambda}_k X^k$. Since $E^{(\ell)}X^k = \binom{k}{\ell} X^{k-l}$ for $k \geq \ell$ we have $E^{(\ell)}\widehat{\varphi}(X) = \sum_{k=\ell}^{\deg \widehat{\varphi}} \widehat{\lambda}_k \binom{k}{\ell} X^{k-l}$, which clearly satisfies $\nu_p(E^{(\ell)}\widehat{\varphi}) \geq 0$.

If $\widehat{\varphi}$ is a polynomial in $X, Y_1, \ldots, Y_d$, by expanding $E^{(\ell)}\widehat{\varphi}$ as above using the product rule (Proposition 5.4), it is sufficient to check that $\nu_p(E^{(\ell)}Y_i) \geq 0$. We will show his by induction on ℓ. By design of the operators D and $E^{(\ell)}$, we see that $E^{(\ell)}\widehat{f}(X, Y_i) = 0$ for all $\ell > 0$. Expanding this relation using the product rule, we find

$$0 = E^{(\ell)}\left(\sum_{k=0}^{d} \widehat{g}_k(X) Y_i^{d-k}\right) = \sum_{k=0}^{d} E^{(\ell)}\left(\widehat{g}_k(X) Y_i^{d-k}\right) = \sum_{k=0}^{d} \sum_{\ell_0+\cdots+\ell_{d-k}=\ell} E^{(\ell_0)}\widehat{g}_k(X) E^{(\ell_1)}Y_i \cdots E^{(\ell_{d-k})}Y_i.$$

By induction and the above discussion, we know that all terms in the above expression lie in the valuation ring of L, except perhaps those including $E^{(\ell)}Y_i$. Hence, we find that

$$\sum_{k=0}^{d} \widehat{g}_k(X) \cdot (d-k) Y_i^{d-k} E^{(\ell)}Y_i \in R_L.$$

In other words, we have $\nu_p(\partial_Y \widehat{f}(X, Y_i) E^{(\ell)}Y_i) \geq 0$. However, note that $\pi(\partial_Y \widehat{f}(X, Y_i)) = \partial_Y f(X, Y_i) \not\equiv 0$, and therefore $\nu_p(\partial_Y \widehat{f}(X, Y_i)) = 0$. Using the additivity of ν_p, we finally at

$$\nu_p(E^{(\ell)}Y_i) = \nu_p(\partial_Y \widehat{f}(X, Y_i) E^{(\ell)}Y_i) - \nu_p(\partial_Y \widehat{f}(X, Y_i)) = \nu_p(\partial_Y \widehat{f}(X, Y_i) E^{(\ell)}Y_i) \geq 0.$$

Finally, we need to check that $\nu_p(E^{(\ell)}\widehat{\varphi}) \geq 0$ for general rational functions. We do this by induction on ℓ again and we pick a polynomial $\widehat{\psi} \in K[X, Y_1, \ldots, Y_d]$ with $\nu_p(\widehat{\psi}) = 0$ and such that $\widehat{\psi}\widehat{\varphi}$ is a polynomial. Applying $E^{(\ell)}$ to $\widehat{\psi}\widehat{\varphi}$ we find

$$E^{(\ell)}(\widehat{\psi}\widehat{\varphi}) = \sum_{j=0}^{\ell} E^{(j)}\widehat{\psi} E^{(\ell-j)}\widehat{\varphi} \in R_L.$$

Note that, by inductive hypothesis, all terms except maybe $\widehat{\psi}E^{(\ell)}\widehat{\varphi}$ are in R_L already. Since $\nu_p(\widehat{\psi}) = 0$, we have $\nu_p(E^{(\ell)}\widehat{\varphi}) = \nu_p(\widehat{\psi}E^{(\ell)}\widehat{\varphi}) - \nu_p(\widehat{\psi}) \geq 0$. This completes the verification of the first condition for $E^{(\ell)}$ to be a well-defined operator on $\mathbb{F}_q(X, Y_1, \ldots, Y_d)$.

The only thing that remains to be checked is that for any two lifts $\widehat{\varphi}_1$ and $\widehat{\varphi}_2$ of $\varphi \in \mathbb{F}_q(X, Y_1, \ldots, Y_d)$ we have $\pi(E^{(\ell)}\widehat{\varphi}_1) = \pi(E^{(\ell)}\widehat{\varphi}_2)$. In fact, it suffices to check that $\nu_p(E^{(\ell)}(\widehat{\varphi}_1 - \widehat{\varphi}_2)) \geq 1$. Since $\pi(\widehat{\varphi}_1) = \pi(\widehat{\varphi}_2)$, we obtain $\nu_p(\widehat{\varphi}_1 - \widehat{\varphi}_2) \geq 1$ and so $\widehat{\varphi}_1 - \widehat{\varphi}_2 = p\widehat{\psi}$ for some rational function $\widehat{\psi} \in L$ with $\nu_p(\widehat{\psi}) \geq 0$. Then

$$\nu_p(E^{(\ell)}(\widehat{\varphi}_1 - \widehat{\varphi}_2)) = \nu_p(pE^{(\ell)}\widehat{\psi}) = 1 + \nu_p(E^{(\ell)}\widehat{\psi}) \geq 1,$$

just as we needed to show. This completes the proof that $E^{(\ell)}$ is well-defined on $\mathbb{F}_q(X, Y_1, \ldots, Y_d)$. □

Now, we arrive at a new definition of the vanishing multiplicity for a polynomial at a given point $x \in \mathbb{F}_q$. We will say that a polynomial $h \in \mathbb{F}_q[X]$ vanishes to order M at x if $E^{(\ell)}h(x) = 0$ for all $0 \leq \ell \leq M-1$. Having made this definition, we need to justify why this complicated operator $E^{(\ell)}$ solves the problem indicated previously, i.e. why this notion is better than the usual notion of the derivative. The following proposition takes care of this point.

Proposition 5.6. If $h(X) \in \mathbb{F}_q[X]$ is a nonzero polynomial for which $E^{(\ell)}h(x) = 0$ for some $x \in \mathbb{F}_q$ and $0 \leq \ell \leq M-1$, then $(X-x)^M | h(X)$.

Proof. This proof is very similar to proof in case of usual polynomials. If we write $h(X) = a_0 + a_1(X-x) + \cdots + a_n(X-x)^n$ and take ℓ-th hyperderivative we obtain

$$E^{(\ell)}h(X) = \sum_{k=\ell}^{n} a_k \binom{k}{\ell} (X-x)^{k-\ell}.$$

From this equation, it is easy to see that $E^{(\ell)}h(x) = 0$ is equivalent to $a_\ell \binom{\ell}{\ell} = 0$. Therefore, the vanishing of $E^{(\ell)}h(x)$ for $0 \leq \ell \leq M-1$ corresponds to the vanishing of the coefficients $a_0, a_1, \ldots, a_{M-1}$, and we obtain $(X-x)^M | h(X)$. □

Corollary 5.7. If $h(X) \in \mathbb{F}_q[X]$ is a nonzero polynomial vanishing to orders $M_1, \ldots, M_k$ at distinct points $x_1, \ldots, x_k$ respectively, then we have $M_1 + \cdots + M_k \leq \deg h(X)$.

Proof. Since h vanishes to order M_i at x_i, Proposition 5.6 implies that $(X-x_i)^{M_i} | h(X)$. Since x_i are distinct, we also have $\prod_{i=1}^{k}(X-x_i)^{M_i} | h(X)$. Since h is nonzero, its degree must be at least as big as $\deg \prod_{i=1}^{k}(X-x_i)^{M_i} = M_1 + \cdots + M_k$, completing the proof. □

Proposition 5.8. For any positive integer m and any $1 \leq \ell < q$ we have $E^{(\ell)}(X^{qm}) = 0$ and $E^{(\ell)}(Y^{qm}) = 0$.

Proof. In fact, one can show that for any rational function φ we have $E^{(\ell)}(\varphi^{qm}) = 0$. Let us begin by taking the lift $\widehat{\varphi}$ of φ and using the definition of $E^{(\ell)}$:

$$E^{(\ell)}\widehat{\varphi}^{qm} = \frac{1}{\ell!} D^{\ell-1}\left(qm\widehat{\varphi}^{qm-1}D\widehat{\varphi}\right) = \frac{qm}{\ell} E^{(\ell-1)}\left(\widehat{\varphi}^{qm-1}D\widehat{\varphi}\right).$$

The first factor projects to zero in $\mathbb{F}_q$ while the second one is still in the valuation ring of L, showing that the projection $\pi(E^{(\ell)}\widehat{\varphi}^{qm}) = E^{(\ell)}\varphi^{qm}$ must be zero. This completes the proof. □

5.2 Step 1: Preprocessing

The main goal of this section is to show that it is not a loss of generality to assume that $f(X,Y)$ is a polynomial of the form $f(X,Y) = Y^d + g_1(X)Y^{d-1} + \cdots + g_d(X)$, where $\deg g_i(X) \leq i$ and $\partial_Y f(X,Y) \not\equiv 0$. For simplicity of notation, we will also define $g_0(X) = 1$. We will also assume f is not linear, as Theorem 5.1 is trivial in this case.

Note that the derivative $\partial_Y f(X,Y)$ is identically zero if and only if all powers of y appearing in $f(X,Y)$ are divisible by p. In other words, this happens if $f(X,Y)$ is a polynomial in x, y^p, where we may write $f(X,Y) = \tilde{f}(X,Y^p)$. Note that the map $y \mapsto y^p$ permutes the elements of $\mathbb{F}_q$ and therefore $\tilde{f}(X,Y)$ and $f(X,Y)$ has the same number of roots in $\mathbb{F}_q{}^2$. In particular, this means that we may perform our analysis on $\tilde{f}$ instead of f, and that the bound (5.1) carries over unchanged. Since this process decreases the degree of Y in f, it terminates and the resulting polynomial does not have all power of y divisible by p.

Now, we may write $f(X,Y) = f_d(X,Y) + f_{d-1}(X,Y) + \cdots + f_0(X,Y)$, where $f_i(X,Y)$ is the homogeneous part of degree i. If we knew that the coefficient of Y^d in $f(X,Y)$ is nonzero, we could scale f by an appropriate constant to bring it into the required form. However, if the coefficient of Y^d is zero, we may consider the polynomial $f_a(X,Y) = f(X+aY,Y)$, where $a \in \mathbb{F}_q$ is a scalar. It is clear that the number of roots of f and f_a is the same. Furthermore, the coefficient of y^d in $f_a(X,Y)$ comes from the highest degree homogeneous part $f_d(X+aY,Y)$. Since this is a homogeneous polynomial, the coefficient next to Y^d can be found by setting $x = 0, y = 1$, meaning that the coefficient of Y^d in $f_a(X,Y)$ is $f_d(a,1)$. Note that $f_d(a,1)$ is a nonzero polynomial of degree d in a, and hence if q is bigger than d there exists a value of a for which $f_d(a,1) \neq 0$.

However, we need to ensure that there are still powers of Y not divisible by p after performing this procedure. In other words, we have to ensure that $\partial_Y f_a(X,Y) \neq 0$. We know from our first step that there is a monomial cX^iY^j in $f(X,Y)$ with $c \neq 0$, j not divisible by p. Then, the coefficient of X^iY^j in $f_a(X,Y)$ will be a nonzero polynomial of degree d in a which has at most d roots. Hence, if q is big enough, we can choose a for which $f_a(X,Y)$ has nonzero coefficients next to both Y^d and X^iY^j, ensuring all the properties we need. Hence, we have proven the following proposition.

Proposition 5.9. For every absolutely irreducible polynomial $f \in \mathbb{F}_q[X,Y]$, there exists an absolutely irreducible polynomial $\tilde{f} \in \mathbb{F}_q[X,Y]$ with the following properties:

- $\tilde{f}(X,Y) = Y^d + g_1(X)Y^{d-1} + \cdots + g_d(X)$, for some polynomials $g_1, \ldots, g_d \in \mathbb{F}_q[X]$ satisfying $\deg g_i \leq i$,
- $\deg \tilde{f} \leq \deg f$,
- $\partial_Y \tilde{f}(X,Y)$ does not vanish identically, and
- equations $\tilde{f}(X,Y) = 0$ and $f(X,Y) = 0$ have the same number of solutions in ${\mathbb{F}_q}^2$.

In the remaining sections, we will always work with $\tilde{f}$ instead of f (although we will, for simplicity of notation, denote it by f).

5.3 Interlude: Variable elimination and degree reduction

The purpose of this section is to introduce two tools that we will use throughout the proof. We begin by presenting a construction that allows us to convert two-variable polynomials into single-variable polynomials, while maintaining information about the zeros of these polynomials on the curve $C/\mathbb{F}_q$. This construction will also be the basis the Step 4, which is presented in the last section of this chapter. However, we prefer to present this equation earlier, since it turns out to be useful in certain arguments of Step 3 as well.

To set up the construction, let $a(X,Y) \in R[X,Y]$ be an arbitrary polynomial, where R is a ring containing $\mathbb{F}_q$.[3] Let us consider the polynomial

$$u(X, Y_1, \ldots, Y_d) = \prod_{i=1}^{d} a(X, Y_i). \tag{5.4}$$

Note that u is a symmetric polynomial in variables $Y_1, \ldots, Y_d$. To simplify notation, we introduce $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ and $\sigma_i(\boldsymbol{Y})$ which denotes the i-th elementary symmetric polynomial in variables $Y_1, \ldots, Y_d$.

By the fundamental theorem of symmetric polynomials, there exists a polynomial $v(X, \sigma_1, \ldots, \sigma_d) \in \mathbb{F}_q[X, \sigma_1, \ldots, \sigma_d]$ which satisfies

$$v(X, \sigma_1(\boldsymbol{Y}), \ldots, \sigma_d(\boldsymbol{Y})) = u(X, Y_1, \ldots, Y_d). \tag{5.5}$$

Then, we define the $f(X,Y)$-*elimination* of $a(X,Y)$ to be the polynomial $b(X)$ defined by

$$b(X) = v(X, -g_1(X), g_2(X), \ldots, (-1)^d g_d(X)). \tag{5.6}$$

We claim that this construction allows us to detect various properties of vanishing of a on $C/\mathbb{F}_q$. We will now try to provide some intuition behind this definition. Suppose $y_1, \ldots, y_d$ are the roots of $f(x,Y)$ for some $x \in \overline{\mathbb{F}_q}$. Then the Vieta formulas guarantee that $\sigma_k(y_1, \ldots y_d) = (-1)^k g_k(x)$, where $g_k(x)$ are the coefficients of the polynomial $f(x,Y) \in \mathbb{F}_q[Y]$. This implies

$$b(x) = v(x, -g_1(x), \ldots, (-1)^d g_d(x)) = u(x, y_1, \ldots, y_d) = \prod_{i=1}^{d} a(x, y_i). \tag{5.7}$$

In other words, b vanishes at a certain point $x \in \overline{\mathbb{F}_q}$ if and only if a vanishes on a point $(x, y_i) \in C/\mathbb{F}_q$. In fact, much more is true - the vanishing multiplicity of b at a point x is at least the sum of the vanishing multiplicities of a over all points (x, y_i). We will postpone the proof of this particular fact until Step 4, and we will describe some simpler properties of this construction.

Proposition 5.10. Let $a \in R[X,Y]$ be a polynomial, and let $b(X)$ be its $f(X,Y)$-elimination with respect to X. Then b is identically zero if and only if $f(X,Y)$ divides $a(X,Y)$.

[3] In our applications, R will either be $\mathbb{F}_q$ or a polynomial ring over $\mathbb{F}_q$.

Proof. Let us begin by assuming $f(X,Y)|a(X,Y)$. Then, for any $x \in \overline{\mathbb{F}_q}$ and any root y_i of $f^{(x)}(Y) = f(x,Y)$, we know $a(x,y_i) = 0$. Using equality (5.7), this gives $b(x) = 0$. Therefore, $b(x)$ vanishes at every point $x \in \mathbb{F}_q$, implying b vanishes identically.

Suppose now that b vanishes identically. Then, for every $x \in \overline{\mathbb{F}_q}$, we have a root y_i of $f^{(x)}(Y)$ for which $a(x,y_i) = 0$. In other words, $f(X,Y)$ and $a(X,Y)$ share infinitely many roots. Proposition 2.29 guarantees this is impossible unless $f(X,Y)$ divides $a(X,Y)$, completing the proof. □

Let us now introduce another useful tool, the degree reduction procedure, which is nothing else than long division of polynomials in a ring. We prefer to state this result as a separate proposition however, because this allows us to refer back to the degree bounds explicitly.

Proposition 5.11. Let $a(X,Y), b(X,Y) \in R[X,Y]$ and suppose $a(X,Y)$ is polynomial of degree d, with a leading term Y^d. Then, there exists a polynomial $c(X,Y) \in R[X,Y]$ such that $b(X,Y) \equiv c(X,Y) \pmod{a(X,Y)}$ and $\deg_Y c(X,Y) \leq d-1, \deg c(X,Y) \leq \deg b(X,Y)$.

Proof. The idea behind the proof is, as stated above, nothing else than long division of polynomials. Namely, if $a(X,Y) = Y^d + \sum_{i=0}^{d-1} Y^i a_{d-i}(X)$, where $\deg a_{d-i}(X) \leq d-i$, we can replace all terms of the form Y^k in $b(X,Y)$ with $k \geq d$ by $-Y^{k-d}\sum_{i=0}^{d-1} Y^i a_{d-i}(X)$, hence reducing the degree of $b(X,Y)$ in Y. Note that such a replacement does not increase the total degree of the polynomial throughout the procedure. Furthermore, note that this replacement does not change the residue class of the polynomial modulo $a(X,Y)$. Hence, we can perform this process until the degree of the remaining polynomial becomes at most $d-1$, at which point we obtain the polynomial $c(X,Y)$. By construction, we have $b(X,Y) \equiv c(X,Y) \pmod{a(X,Y)}$ and since our process did not increase the total degree of the polynomial, we have $\deg c(X,Y) \leq \deg b(X,Y)$. □

5.4 Step 2: Splitting the roots

In this section, we begin to set up the proof of Theorem 5.1. The way we will count the roots of $f(X,Y)$ is by fixing $X = x$ and analyzing the roots of the univariate polynomial in Y, $f^{(x)}(Y) = f(x,Y)$.

We begin by a simple observation - for almost all values of x, the polynomial $f^{(x)}(Y)$ has d distinct roots in $\overline{\mathbb{F}_q}$.

Proposition 5.12. Let $\mathfrak{S}$ be the set of $x \in \mathbb{F}_q$ for which $f^{(x)}(Y) = f(x,Y) = 0$ has d distinct roots in Y. Then $|\mathfrak{S}| \geq q - 2d^2$.

Proof. Let $\Delta_{f^{(x)}}(x)$ be the discriminant of the polynomial $f^{(x)}(Y)$. The key property we will use is that a polynomial has d distinct roots if and only if its discriminant is nonzero. The coefficients of the polynomial $f^{(x)}(Y)$ are themselves polynomials in x, and hence the discriminant of $f^{(x)}$ is a polynomial in x. The discriminant $\Delta_{f^{(x)}}$ is a polynomial of degree $2d-d$ in the coefficients of $f^{(x)}$ and hence of degree at most $2(d-1)^2$ in x. Hence, $\Delta_{f^{(x)}}$ has at most $2(d-1)^2$ roots (note that $\Delta_{f^{(x)}}$ is not identically zero, since $f(x,y)$ is absolutely irreducible). The conclusion is that $\mathfrak{S}$ contain all but at most $2(d-1)^2 < 2d^2$ elements of $\mathbb{F}_q$. □

Here, as throughout the rest of the proof, we will think of d as constant as q as going to infinity. In this setting, $\mathfrak{S}$ is really almost all of $\mathbb{F}_q$. For every $x \in \mathfrak{S}$, we set

$$I_1(x) = \{y \in \mathbb{F}_q | f(x,y) = 0\} \qquad I_2 = \{y \in \overline{\mathbb{F}_q} \backslash \mathbb{F}_q | f(x,y) = 0\}.$$

Our goal will be to bound the sum of sizes of $I_1(x)$ over all $x \in \Sigma$, both from above and from below. However, since we are able to establish only the upper bounds using the polynomial method, the lower bounds on $I_1(x)$ will come from the upper bounds on $I_2(x)$, since for all $x \in \mathfrak{S}$ we have $|I_1(x)| + |I_2(x)| = d$. Let us now state the precise bounds we obtain. These bounds will be proven at the end of Step 4.

Proposition 5.13. In the setup described above, we have

$$\sum_{x \in \mathfrak{S}} |I_1(x)| \leq q + O(q^{1/2}d^3) \text{ and } \sum_{x \in \mathfrak{S}} |I_2(x)| \leq (d-1)q + O(q^{1/2}d^3). \tag{5.8}$$

This is a good place to introduce a minor notational convention which we will follow throughout the rest of the proof. Since many arguments are very similar for $I_1(x)$ and $I_2(x)$, we will often phrase the bounds in terms of $I_t(x)$, where the index t is either 1 or 2.

Now, we will demonstrate how Proposition 5.13 lead to the proof of Theorem 5.1.

Proof of Theorem 5.1 assuming Proposition 5.13. Let N denote the number of solution to $f(x,y)=0$ in ${\mathbb{F}_q}^2$. The upper bound on N can be established from the observation that for every x outside $\mathfrak{S}$, $f^{(x)}(Y)$ has at most d roots, and therefore

$$N \leq d|\mathbb{F}_q \backslash \mathfrak{S}| + \sum_{x\in\mathfrak{S}} |I_1(x)| \leq 2d^3 + q + O(q^{1/2}d^3) \leq q + O(q^{1/2}d^3).$$

The lower bound follows from the bounds on $|I_2(x)|$, by noting that $|I_1(x)| + |I_2(x)| = d$ for all $x \in \mathfrak{S}$:

$$N \geq \sum_{x\in\mathfrak{S}} |I_1(x)| = d|\mathfrak{S}| - \sum_{x\in\mathfrak{S}} |I_2(x)| \geq dq - 2d^3 - (d-1)q - O(q^{1/2}d^3) \geq q - O(q^{1/2}d^3).$$

□

The main tool we will use in bounding the size of I_t will be a certain polynomial $e_t(X,Y,Y^q)$ which will vanish on all points (x,y) where $x \in \Sigma, y \in I_t(x)$. Then, the idea will be to bound the number of common zeros of e_t and f.

Proposition 5.14. Define the polynomials $e_1(X,Y,Y') = Y' - Y$ and

$$e_2(X,Y,Y') = \sum_{i=0}^{d-1} g_i(X)({Y'}^{d-1} + {Y'}^{d-2}Y + \cdots + Y^{d-1}).$$

Then $e_t(x,y,y^q) = 0$ for all $x \in \mathfrak{S}, y \in I_t(x)$.

Proof. The case $t=1$ is obvious, since $y^q = y$ for all $y \in I_1(x) \subset \mathbb{F}_q$ and therefore $e_t(x,y,y^q) = 0$. If $y \in I_2(x)$, we have that $f(x,y) = 0$ and so $0 = f(x,y)^q = f(x^q, y^q) = f(x,y^q)$. In particular, we have $f(x,y) = f(x,y^q)$ and therefore

$$0 = f(x,y^q) - f(x,y) = \sum_{i=0}^{d} g_i(x)y^{q(d-i)} - \sum_{i=0}^{d} g_i(x)y^{d-i} = \sum_{i=0}^{d-1} g_i(x)(y^q - y)\left(y^{d-i-1} + y^q y^{d-i-2} + \cdots + y^{q(d-i)}\right).$$

Since $y \in I_2(x)$ is not in $\mathbb{F}_q$, we know $y^q - y \neq 0$ and the conclusion is $e_2(x,y,y^q) = 0$, as claimed. □

If we were to apply Bezout's theorem directly to polynomials $e_1(X,Y,Y^q) = Y^q - Y$ and $f(X,Y)$, one could bound the number of common points by qd. However, this is very far from the bound in Proposition 5.13. In a way, the key to the proof of Theorem 5.1 lies precisely in improving this bound, by using the polynomial method.

From now on, we will denote by d_t the degree of $e_t(X,Y,Y')$ in variable Y'. In other words, we have $d_1 = 1, d_2 = d-1$.

5.5 Step 3: Constructing the auxiliary polynomial

In this section, we will construct two auxiliary polynomials in variables X,Y, which we call $c_t(X,Y)$ for $t \in \{0,1\}$. These polynomials should be of low degree, vanish at all points (x,y) for $x \in \mathfrak{S}, y \in I_t(x)$ and not be divisible by f. In the following proposition, we present the precise bounds required by these polynomials. From now on, we will assume that M is a parameter satisfying $d|M$ and $d^2 < M < \frac{q^{1/2}}{2d}$, whose value we will optimize at the end.

Proposition 5.15. There exist polynomials $c_t(X,Y)$, for $t \in \{1,2\}$, satisfying the following three properties

- $f(X,Y)$ does not divide $c_t(X,Y)$,

- $E^{(\ell)}c_t(x,y) = 0$ for all $x \in \mathfrak{S}, y \in I_t(x), 0 \le \ell \le M-1$. In other words, the polynomial c_t vanishes at (x,y) to order M.
- The degree of c_t is bounded by $\deg c_t \le \frac{d_t}{d} qM + 2qd$.

We will look for c_t in the form of $c_t(X,Y) = (\partial_Y f(X,Y))^{2M} a_t(X,Y)$. The reason this is useful is that the definition of D (and consequently $E^{(\ell)}$) involved dividing by $\partial_Y f$, which may introduce denominators in our polynomial. However, it is much easier for us if we keep the derivatives of c_t to be polynomials, especially because we do not have to pay much for this convenience in increasing the degree. Namely, multiplying a polynomial by $(\partial_Y f)^{2M}$ increases its degree by at most $2Md \le q^{1/2}$, which is much smaller than the leading term of $\deg c_t \sim \frac{d_t}{d} qM$. Let us now justify this intuition, and show that all hyperderivatives of $c_t(X,Y)$ are indeed polynomials, if $c_t(X,Y)$ is chosen to be of this form.

Proposition 5.16. Let $a_t(X,Y) \in \mathbb{F}_q[X,Y]$ be a polynomial and define $c_t(X,Y) = (\partial_Y f(X,Y))^{2M} a_t(X,Y)$. Then, for all $1 \le \ell \le M$, we have

$$E^{(\ell)}c_t(X,Y) = (\partial_Y f(X,Y))^{2(M-\ell)} a_t^{(\ell)}(X,Y),$$

where $a_t^{(\ell)}(X,Y) \in \mathbb{F}_q[X,Y]$ is a polynomial of degree at most $\deg a_t^{(\ell)} \le \deg a_t + 2d\ell$.

Proof. Let $\widehat{a}_t(X,Y), \widehat{f}(X,Y)$ be the lifts of $a_t(X,Y), f(X,Y)$, chosen in the valuation ring of ν_p on $K[X,Y]$. If we define $\widehat{a}_t^{(\ell)}(X,Y)$ through the relation $E^{(\ell)}\left((\partial_Y \widehat{f}(X,Y))^{2M}\widehat{a}_t(X,Y)\right) = (\partial_Y \widehat{f}(X,Y))^{2(M-\ell)}\widehat{a}_t^{(\ell)}(X,Y)$. We will show by induction on ℓ that $\widehat{a}_t^{(\ell)}(X,Y)$ is a polynomial of degree $\le \deg \widehat{a}_t + 2d\ell$. We have

$$\begin{aligned}
E^{(\ell+1)}\left((\partial_Y \widehat{f})^{2M}\widehat{a}_t\right) &= \frac{1}{\ell+1} DE^{(\ell)}\left((\partial_Y \widehat{f})^{2M}\widehat{a}_t\right) \\
&= \frac{1}{\ell+1} D\left((\partial_Y \widehat{f})^{2M-2\ell}\widehat{a}_t^{(\ell)}\right) \\
&= \frac{(2M-2\ell)\partial_Y \widehat{f}^{2M-2\ell-1}\left(\partial_{XY}\widehat{f} - \frac{\partial_X \widehat{f}}{\partial_Y \widehat{f}}\partial_{YY}\widehat{f}\right)\widehat{a}_t^{(\ell)} + \partial_Y \widehat{f}^{2M-2\ell}\left(\partial_X \widehat{a}_t^{(\ell)} - \frac{\partial_X \widehat{f}}{\partial_Y \widehat{f}}\partial_Y \widehat{a}_t^{(\ell)}\right)}{\ell+1} \\
&= \frac{\partial_Y \widehat{f}^{2M-2\ell-2}}{l+1}\left(\left(\partial_{XY}\widehat{f}\partial_Y \widehat{f} - \partial_X \widehat{f}\partial_{YY}\widehat{f}\right)\widehat{a}_t^{(\ell)} + \partial_Y \widehat{f}\left(\partial_X \widehat{a}_t^{(\ell)}\partial_Y \widehat{f} - \partial_X \widehat{f}\partial_Y \widehat{a}_t^{(\ell)}\right)\right).
\end{aligned}$$

Since $\deg \partial_Y \widehat{f}, \deg \partial_X \widehat{f} \le d$, we easily find that $\deg \widehat{a}_t^{(\ell+1)} \le \deg \widehat{a}_t^{(\ell)} + 2d$, thus giving us $\deg \widehat{a}_t^{(\ell+1)} \le \deg a_t + 2d\ell$ as needed.

Furthermore, we have $\nu_p(\widehat{a}_t^{(\ell)}) \ge 0$. To show this recall that we have $\nu_p(\partial_Y \widehat{f}^{2M-2\ell}\widehat{a}_t^{(\ell)}) = \nu_p(E^{(\ell)}(\partial_Y \widehat{f}^{2M}\widehat{a}_t)) \ge 0$ and $\nu_p(\partial_Y \widehat{f}) = 0$ (since $\pi(\partial_Y \widehat{f}) = \partial_Y f \ne 0$). Thus, we have

$$0 \le \nu_p(\partial_Y \widehat{f}^{2M-2\ell}\widehat{a}_t^{(\ell)}) = (2M-2l)\nu_p(\partial_Y \widehat{f}) + \nu_p(\widehat{a}_t^{(\ell)}) = \nu_p(\widehat{a}_t^{(\ell)}).$$

Hence, we can take the projection $\pi(\widehat{a}_t^{(\ell)}) = a_t^{(\ell)} \in \mathbb{F}_q[X,Y]$ and we easily find the required degree constraints $\deg a_t^{(\ell)} \le \deg \widehat{a}_t^{(\ell)} \le \deg a_t + 2d\ell$. This completes the proof. □

Now, the question reduces to constructing $a_t(X,Y)$ such that $a_t^{(\ell)}(x,y) = 0$ for all $x \in \mathfrak{S}, y \in I_t(x)$, or alternatively $a_t^{(\ell)}(X,Y) \in (f(X,Y), e_t(X,Y,Y^q))$. Note that the coefficients of $a_t^{(\ell)}(X,Y)$ are linear in the coefficients of $a_t(X,Y)$, simply because the operators D^ℓ and $E^{(\ell)}$ are linear. The idea behind constructing a polynomial $a_t(X,Y)$ we need is to pick it in the form

$$a_t(X,Y) = \sum_{j=0}^{K}\sum_{k=0}^{d-1}\sum_{i=0}^{d-1} a_{ijk}(X) X^{qj} Y^{i+qk}, \tag{5.9}$$

where $a_{ijk}(X)$ will be of degree at most $\deg a_{ijk}(X) \le \frac{q}{d} - 3d - j$ and $K = \frac{d_t}{d}M + d$.

The advantage of choosing $a_t(X,Y)$ in this form is that as long as at least one of the coefficients $a_{ijk}(X)$ is nonzero, we can always guarantee that $f(X,Y)$ does not divide $a_t(X,Y)$, as the following proposition shows.

Proposition 5.17. Let $F(X,Y,Z,W) \in \mathbb{F}_q[X,Y,Z,W]$ be a nonzero polynomial with $\deg_X F \le \frac{q}{d} - d, \deg_Y F \le d-1$ and $\deg_W F \le d-1$. Then, $f(X,Y)$ does not divide $F(X,Y,X^q,Y^q)$.

Proof. Let us begin by considering the polynomial $G(X,Y,Z)$ which is the $f(Z,W)$-elimination of the polynomial $F(X,Y,Z,W)$.[4] By construction, we have $\deg_{X,Y} G(X,Y,Z) \leq d \deg_{X,Y} F(X,Y,Z,W) < q$. Furthermore, we take $H(X,Y,Z) \in \mathbb{F}_q[X,Y,Z]$ such that $H(X,Y,Z) \equiv G(X,Y,Z) \pmod{f(X,Y)}$ and $\deg_Y H(X,Y,Z) \leq d-1$, $\deg_{X,Y} H(X,Y,Z) \leq \deg_{X,Y} G(X,Y,Z) < q$ (the existence of such H is guaranteed by Proposition 5.11).

The first step of our proof will be to show that if $f(X,Y) | F(X,Y,X^q,Y^q)$, then we also have $H(X,Y,X^q) = 0$. Then, we will use the degree conditions on H to show that $H(X,Y,Z) = 0$, which will suffice to show that $F(X,Y,Z,W) \in (f(X,Y), f(Z,W))$ by a simple application of Proposition 5.10. Finally, we will obtain a contradiction from the degree conditions on $F(X,Y,Z,W)$.

Now, we execute the first step of the proof. We fix $x \in \overline{\mathbb{F}_q}$ and let $y_1, \ldots, y_d$ be the roots of $f^{(x)}(Y) = f(x,Y)$. Then $y_1^q, \ldots, y_d^q$ are roots of $f(x^q,Y)$ (since $0 = f(x,y_i)^q = f(x^q,y_i^q)$) the Vieta formulas ensure that $\sigma_i(y_1^q, \ldots, y_d^q) = (-1)^i g_i(x^q)$. Since $f(X,Y) | F(X,Y,X^q,Y^q)$ we have $F(x,y_i,x^q,y_i^q) = 0$ and therefore $G(x,y_i,x^q) = 0$ by equality (5.7). Since $f(x,y_i) = 0$, we also have $H(x,y_i,x^q) = 0$. If x was such that $f^{(x)}(Y)$ has d distinct roots, the polynomial $H(x,Y,x^q)$ would also have d roots. Since $\deg_Y H(x,Y,x^q) \leq d-1$ we must have $H(x,Y,x^q) \equiv 0$. But this holds for almost any $x \in \overline{\mathbb{F}_q}$ (as shown by Proposition 5.12), implying $H(X,Y,X^q) = 0$, as claimed.

Now, we want to show $H(X,Y,Z) = 0$. Recall that $\deg_X H \leq q-1, \deg_Y H \leq d-1$ and therefore we can write $H(X,Y,Z) = \sum_{i=0}^{q-1} \sum_{j=0}^{d-1} \sum_{k \geq 0} h_{ijk} X^i Y^j Z^k$. Plugging in $Z = X^q$ we have

$$H(X,Y,X^q) = \sum_{i=0}^{q-1} \sum_{j=0}^{d-1} \sum_{k \geq 0} h_{ijk} X^{i+kq} Y^j,$$

where the key observation is that all monomials $X^{i+qk}Y^j$ are different and therefore the only way to make $H(X,Y,X^q) = 0$ would be to have $h_{ijk} = 0$ for all i,j,k, meaning $H(X,Y,Z) = 0$.

Finally, this means that $f(X,Y) | G(X,Y,Z)$ and therefore $\widetilde{G}(X,Y,Z) = 0$, where $\widetilde{G}$ is the residue of G modulo f, i.e. $\widetilde{G}(X,Y,Z) \in \mathbb{F}_q[X,Y,Z,W]/(f(X,Y))$. If $\widetilde{F}(X,Y,Z,T)$ denotes the residue of $F(X,Y,Z,W)$, note that $\widetilde{G}(X,Y,Z)$ is still a $\widetilde{f}(Z,W)$-elimination of $\widetilde{F}(X,Y,Z,W)$. In particular, since $\widetilde{G}(X,Y,Z) = 0$, we must have $\widetilde{f}(Z,W) | \widetilde{F}(X,Y,Z,W)$. But this is just another way to say that $F(X,Y,Z,W)$ belongs to the ideal $I = (f(X,Y), f(Z,W))$.

Now, we claim that the monomials $Y^i W^j$ for $0 \leq i,j \leq d-1$ are $\mathbb{F}_q[X,Z]$-linearly independent in $\mathbb{F}_q[X,Y,Z,W]/I$, and that therefore $F(X,Y,Z,W)$ cannot be in this ideal unless it is equal to zero. Suppose for the contrary, that we had an relation of the form

$$F(X,Y,Z,W) = \sum_{i=0}^{d-1} \sum_{j=0}^{d-1} p_{ij}(X,Z) Y^i W^j = \alpha(X,Y,Z,W) f(X,Y) + \beta(X,Y,Z,W) f(Z,W).$$

Choosing a root (z,w) of $f(Z,W) = 0$, this relation implies that $f(X,Y) | \sum_{i=0}^{d-1} Y^i \sum_{j=0}^{d-1} p_{ij}(X,z) w^j$. Since the degree in Y of the sum is at most $d-1$, this is impossible unless all its coefficients are zero. In other words, we have $\sum_{j=0}^{d-1} p_{ij}(X,z) w^j = 0$ for all i. If $f(z,W)$ has d distinct roots, so does the polynomial $\sum_{j=0}^{d-1} p_{ij}(x,z) W^j$ in W, for any x, and thus we must have $p_{ij}(x,z) = 0$ for any $x \in \overline{\mathbb{F}_q}$ and almost any $z \in \overline{\mathbb{F}_q}$. But this means that all coefficients $p_{ij}(X,Z)$ are zero, giving a contradiction. This completes the proof. □

Now, we will show how to pick the coefficients of $a_{ijk}(X,Y)$ such that $a_t^{(\ell)}(X,Y)$ vanishes on all pairs (x,y) with $x \in \mathfrak{S}, y \in I_t(x)$. The idea is that, for each $0 \leq \ell \leq M-1$, we construct another polynomial $b_t^{(\ell)}(x,y,y^q)$ of much smaller degree than $a_t^{(\ell)}$ such that $a_t^{(\ell)}$ vanishes on all pairs (x,y) with $x \in \mathfrak{S}, y \in I_t(x)$ if the polynomial $b_t^{(\ell)}(X,Y,Y^q)$ vanishes on these pairs too.

Proposition 5.18. There exists a polynomial $b_t^{(\ell)}(X,Y,Y') \in \mathbb{F}_q[X,Y,Y']$ satisfying the following degree bounds

$$\deg_X b_t^{(\ell)} \leq \frac{q}{d} + 2d\ell \qquad \deg_Y b_t^{(\ell)} \leq d-1 \qquad \deg_{Y'} b_t^{(\ell)} \leq d_t - 1$$

[4] In more concrete terms, we define the symmetric polynomial $u(X,Y,Z,W_1,\ldots,W_d) = \prod_{i=1}^n F(X,Y,Z,W_i)$ and find a polynomial v for which $u(X,Y,Z,\boldsymbol{W}) = v(X,Y,Z,\sigma_1(\boldsymbol{W}),\ldots,\sigma_d(\boldsymbol{W}))$. Then, we define $G(X,Y,Z) = v(X,Y,Z,-g_1(Z),\ldots,(-1)^d g_d(Z))$. This corresponds to our construction from Section 5.3 with $R = \overline{\mathbb{F}_q}[X,Y]$ and the variables Z,W.

and the following property: if $b_t^{(\ell)}(X,Y,Y^q)$ vanishes on all points (x,y) with $x \in \mathfrak{S}, y \in I_t(x)$, then $a_t^{(\ell)}(X,Y)$ vanishes on all these points too. Moreover, $b_t^{(\ell)}$ can be chosen to that its coefficients are linear combinations of the coefficients of $a_t(X,Y)$.

Proof. The main ingredient of this proof is the degree reduction procedure described in Proposition 5.11. We will perform two basic degree reduction steps to get from $a_t^{(\ell)}$ to $b_t^{(\ell)}$. In the first step, we will reduce the degree of Y' using that $e_t(x,y,y^q)$ vanishes for $x \in \mathfrak{S}, y \in I_t(x)$. Then, we will use the fact that f vanishes at all of these points to reduce the degree of Y to $d-1$, which will essentially complete the proof. Let us now present the details.

Since the statement of this proposition concerns the vanishing of $a_t^{(\ell)}(X,Y)$, let us begin by calculating $a_t^{(\ell)}(X,Y)$. If introduce the notation $a_{jk}(X,Y) = \sum_{i=0}^{d-1} Y^i a_{ijk}(X)$, we have

$$a_t(X,Y) = \sum_{j=0}^{K} \sum_{k=0}^{d} a_{jk}(X,Y) X^{qj} Y^{qk}.$$

Proposition 5.8 implies that $E^{(\ell)}(X^{qj}) = E^{(\ell)}(Y^{qk}) = 0$ since $\ell < M < q$. Hence, we may write

$$a_t^{(\ell)}(X,Y) = \sum_{j=0}^{K} \sum_{k=0}^{d-1} a_{jk}^{(\ell)}(X,Y) X^{qj} Y^{qk}, \tag{5.10}$$

where $a_{jk}^{(\ell)}(X,Y)$ are defined through the equation $E^{(\ell)}\left((\partial_Y f)^{2M} a_{jk}^{(\ell)}\right) = (\partial_Y f)^{2(M-\ell)} a_{jk}^{(\ell)}$ and therefore satisfy $\deg a_{jk}^{(\ell)} \le \deg a_{jk} + 2d\ell \le \frac{q}{d} - 2d - j + 2d\ell$.

Furthermore, since $b_t^{(\ell)}$ is a polynomial in variables X, Y, Y' (where Y' corresponds to Y^q in the end), it will be easiest if we considered a polynomial $a_t^{(\ell)}$ as a polynomial in three variables

$$a_t^{(\ell)}(X,Y,Y') = \sum_{j=0}^{K} \sum_{k=0}^{d-1} a_{jk}^{(\ell)}(X,Y) X^{qj} Y'^{k}.$$

In our first degree reduction step, we can reduce the degree of $a_t^{(\ell)}(X,Y,Y')$ modulo $e_t(X,Y,Y')$ to obtain a polynomial $\tilde{b}_t^{(\ell)}(X,Y,Y') \in \mathbb{F}_q[X,Y,Y']$ with $\deg \tilde{b}_t^{(\ell)} \le \frac{q}{d} + 2d\ell$ and $\deg_{Y'} \tilde{b}_t^{(\ell)}(X,Y,Y') = 0 = d_t - 1$. In particular, we have

$$a_t^{(\ell)}(X,Y,Y^q) \equiv \tilde{b}_t^{(\ell)}(X,Y,Y^q) \pmod{e_t(X,Y,Y^q)}.$$

The polynomial $\tilde{b}_t^{(\ell)}(X,Y,Y')$ satisfies all of the degree constraints from the statement of this proposition except the condition on $\deg_Y b_t^{(\ell)}(X,Y,Y')$. Hence, we need to perform another degree reduction step, this time modulo $f(X,Y)$. Namely, we can find a polynomial $b_t^{(\ell)}(X,Y,Y')$ which satisfies

$$b_t^{(\ell)}(X,Y,Y') \equiv \tilde{b}_t^{(\ell)}(X,Y,Y') \pmod{f(X,Y)},$$

and the degree constrains $\deg_X b_t^{(\ell)} \le \deg_{X,Y} \tilde{b}_t^{(\ell)}, \deg_Y b_t^{(\ell)} \le d-1$ and $\deg_{Y'} b_t^{(\ell)}(X,Y,Y') \le d_t - 1$.

Observe now that our construction yields $a_t^{(\ell)}(X,Y,Y^q) \equiv b_t^{(\ell)}(X,Y,Y^q) \pmod{(f(X,Y), e_t(X,Y,Y^q))}$, and that both $f(X,Y)$ and $e_t(X,Y,Y^q)$ vanish for all $x \in \mathfrak{S}, y \in I_t(x)$. This means that $a_t^{(\ell)}(X,Y,Y^q)$ and $b_t^{(\ell)}(X,Y,Y^q)$ take the same values on these points, just as required. □

Now, we are ready to construct our auxiliary polynomial $a_t(X,Y)$ and prove Proposition 5.15.

Proof of Proposition 5.15. As suggested in the informal overview, we look for the polynomial $c_t(X,Y)$ in the form $c_t(X,Y) = (\partial_Y f(X,Y))^{2M} a_t(X,Y)$, and for $a_t(X,Y)$ in the form given by (5.9). The fact that $f(X,Y)$ does not divide $c_t(X,Y)$ is ensured by Proposition 5.17 and the fact that $\partial_Y f(X,Y)$ is coprime to $f(X,Y)$ (since $f(X,Y)$ is absolutely irreducible). Furthermore, it is not hard to observe from (5.9) that $\deg a_t \le qK + q(d-1) + \frac{q}{d} \le \frac{d_t}{d} Mq + 2qd$, as claimed.

Finally, we need to ensure that $a_t^{(\ell)}(x,y)$ vanished for all $x \in \mathfrak{S}, y \in I_t(x)$. Performing the degree reduction procedure, Proposition 5.18, for any $0 \le \ell \le M-1$ we obtain a polynomial $b_t^{(\ell)}(X,Y,Y')$ satisfying $a_t^{(\ell)}(x,y) = b_t^{(\ell)}(x,y,y^q)$ on the points $x \in \mathfrak{S}, y \in I_t(x)$. In fact, it suffices to ensure that all coefficients of $b_t^{(\ell)}$ vanish,

for $0 \le \ell \le M-1$. Noting that the coefficients of the polynomials $b_t^{(\ell)}$ depend linearly on the coefficients of $a_t(X,Y)$ basic linear algebra and dimension counting are sufficient to complete the proof.

The number of available coefficients for $a_t(x,y)$ is at least

$$A \ge \sum_{j=0}^{K}\sum_{k=0}^{d-1}\sum_{i=0}^{d-1} \deg a_{ijk} \ge d^2 \sum_{j=0}^{K} \frac{q}{d} - 3d - j = qKd - 3d^3K - \binom{K+1}{2}$$
$$\ge d_t qM + d^2 q - 3d^3 M - 3d^4 - 2M^2 \ge d_t qM + \frac{d^2 q}{2},$$

if q is big enough compared to d and $M \le q^{1/2}$. On the other hand, the number of linear constraints on these coefficients equals the number of coefficients of all polynomials $b_t^{(\ell)}$ for $0 \le \ell \le M-1$, which is

$$B = \sum_{\ell=0}^{M-1} dd_t \left(\frac{q}{d} + 2d\ell + 1\right) \le d_t qM + 3d^3 M.$$

Again, as long as q is big enough compared to d, we have $A > B$, thus ensuring that there exists a nonzero assignment of $a_{ijk}(x)$ satisfying that all polynomials $b_t^{(\ell)}$ are identically zero. This guarantees that $c_t(x,y)$ vanishes for all $x \in \mathfrak{S}, y \in I_t(x)$ and the proof is thus complete. □

5.6 Step 4: Eliminating variables

In this section, we show how to go from a bivariate auxiliary polynomial $c_t(X,Y)$ to a univariate auxiliary polynomial $h_t(X)$ using the technique from Section 5.3. Before we do this, let us explain why this is needed.

One way we could try to approach this completion of the proof is to bound the sum of the vanishing multiplicities of a bivariate polynomial by its degree using the Schwartz-Zippel lemma. A version of the Schwartz-Zippel lemma states that the sum of vanishing multiplicities of a bivariate polynomial $c_t(X,Y) \in \mathbb{F}_q[X,Y]$ is at most nq. Here, the vanishing multiplicity of the polynomial $c_t(X,Y)$ at the point $(x,y) \in \overline{\mathbb{F}_q}^2$ is defined as the largest integer M such that all partial derivatives of $c_t(X,Y)$ of order $0,1,\dots,M-1$ vanish at (x,y). However, there are several problems with applying this lemma in our case.

First of all, the definition of the vanishing multiplicity used in the Schwartz-Zippel lemma is different from the version used in this proof. In this proof, the vanishing multiplicity of $c_t(X,Y)$ is inherently a 1-dimensional quantity since the definition of D was tailored to imitate taking the derivative along a curve $f(X,Y) = 0$. Hence, the Schwartz-Zippel lemma does not really apply here.

Secondly, even if we somehow managed to apply this lemma, it would give a bound on the number of zeros of f of the order $\deg c_t \cdot q \sim \frac{d_t}{d} q^2 M$, which will be useless in our application. Hence, in order to bound the sum of vanishing multiplicities for single variable by using Proposition 5.6, we need to construct a univariate polynomial out of $c_t(X,Y)$. The precise way we eliminate the variable Y differs slightly from Schmidt's approach in [25], which uses field extensions and norms of elements.

Throughout this section, we will make a careful distinction about equality of polynomials and equality of their values. Therefore, throughout this section, we remind the reader that $X, Y_1, \dots, Y_d$ denote the variables and $x, y_1 \dots, y_d \in \overline{\mathbb{F}_q}$ denote their values.

We construct $h_t(X)$ as a $f(X,Y)$-elimination of $c_t(X,Y)$. As explained in Section 5.3, this construction guarantees that for any $x \in \mathfrak{S}$ and $y_1, \dots, y_d$ roots of $f^{(x)}$ one has

$$h_t(x) = \prod_{i=1}^{d} c_t(x, y_i). \tag{5.11}$$

The key feature of this definition is that we also have the following equality of derivatives

$$Dh_t(x) = \sum_{i=1}^{d} Dc_t(x,y_i) \prod_{j \ne i} c_t(x,y_j) = Du(x, y_1, \dots, y_d),$$

where $u(X, Y_1, \dots, Y_d) = \prod_{i=1}^{d} c_t(X, Y_i)$. This equality guarantees that h_t has a double zero at x if the polynomial c_t has a double zero at all of y_i, for example. This will in turn allow us to transfer the vanishing multiplicities from c_t to h_t as we intended.

Note that the equality of the derivatives does not follow directly from (5.7) or (5.11), since these expressions only talk about the equality of elements in $\overline{\mathbb{F}_q}$. Hence, a little effort will be needed to derive the equality of derivatives. The following lemma is actually a little big more general than this, primarily because we will reuse it later on to show the equality of hyperderivatives as well.

Proposition 5.19. Let $x \in \mathfrak{S}$ and $y_1, \dots, y_d$ be the roots of $f(x, Y)$. Let $u(X, Y_1, \dots, Y_d)$ be any $\mathbb{F}_q$-polynomial symmetric in $Y_1, \dots, Y_d$, and let v be the polynomial satisfying

$$v(X, \sigma_1(\boldsymbol{Y}), \dots, \sigma_d(\boldsymbol{Y})) = u(X, Y_1, \dots, Y_d).$$

If we define $h(X) = v_t(X, -g_1(X), g_2(X), \dots, (-1)^d g_d(X))$, we have

$$Dh_t(x) = Du(x, y_1, \dots, y_d).$$

Proof. The proof of this statement mainly relies on unpacking the definitions and showing a simple statement about symmetric polynomials. Let us fix some notation - we denote the partial derivative along $(i+1)$-st argument of u, v by $\partial_i u, \partial_i v$ for $i \in \{0, \dots, d\}$. Then, we have

$$Dh(X) = \partial_0 v(X, -g_1(X), \dots, (-1)^d g_d(X)) + \sum_{k=1}^{d} (-1)^k \partial_X g_k(X) \partial_k v(X, -g_1(X), \dots, (-1)^d g_d(X))$$

and

$$Du(X, Y_1, \dots, Y_d) = \partial_0 u(X, Y_1, \dots, Y_d) + \sum_{i=1}^{d} \frac{\partial_X f(X, Y_i)}{\partial_Y f(X, Y_i)} \partial_i u(X, Y_1, \dots, Y_d).$$

If we differentiate the equality $u(X, Y_1, \dots, Y_d) = v(X, \sigma_1(\boldsymbol{Y}), \dots, \sigma_d(\boldsymbol{Y}))$ with respect to Y_i we obtain

$$\partial_i u(X, Y_1, \dots, Y_d) = \sum_{k=1}^{d} \partial_k v(X, \sigma_1(\boldsymbol{Y}), \dots, \sigma_d(\boldsymbol{Y})) \partial_{Y_i} \sigma_k(\boldsymbol{Y}).$$

It is well-known that the derivative of the k-th symmetric polynomial $\sigma_k(\boldsymbol{Y})$ with respect to y_i is the $(k-1)$-th symmetric polynomial $\sigma_{k-1}(\boldsymbol{Y}_i^-)$, where $\boldsymbol{Y}_i^-$ denotes the $(d-1)$-tuple obtained from $\boldsymbol{Y}$ by removing Y_i. Therefore, we have

$$Du(X, Y_1, \dots, Y_d) = \partial_0 v(X, Y_1, \dots, Y_d) + \sum_{k=1}^{d} \partial_k v(X, Y_1, \dots, Y_d) \sum_{i=1}^{d} \frac{\partial_X f(X, Y_i)}{\partial_Y f(X, Y_i)} \sigma_k(\boldsymbol{Y}_i^-).$$

Plugging in $f(X, Y) = \sum_{j=0}^{d} g_j(X) Y^{d-j}$, we have

$$Du(X, Y_1, \dots, Y_d) = \partial_0 v(X, Y_1, \dots, Y_d) + \sum_{k=1}^{d} \sum_{j=1}^{d} \partial_k v(X, Y_1, \dots, Y_d) Dg_j(X) \sum_{i=1}^{d} \frac{Y_i^{d-j}}{\partial_Y f(X, Y_i)} \sigma_{k-1}(\boldsymbol{Y}_i^-)$$

Hence, in order to show $Dh_t(x) = Du(x, y_1, \dots, y_d)$, we only need to check

$$(-1)^k \mathbf{1}_{k=j} = \sum_{i=1}^{d} \frac{y_i^{d-j}}{\partial_Y f(x, y_i)} \sigma_{k-1}(\overline{\boldsymbol{Y}}_i^-). \tag{5.12}$$

We claim that relation (5.12) follows directly from the formula for the inverse of the Vandermonde matrix. Namely, let us define the $d \times d$ matrix A by setting $A_{ji} = y_i^{d-j}$. It is well known that the entries of the inverse A^{-1} are given by

$$(A^{-1})_{ik} = \frac{(-1)^k \sigma_{k-1}(\overline{\boldsymbol{Y}}_i^-)}{\prod_{m \neq i} (y_i - y_m)}.$$

Noting that $\prod_{m \neq i} (y_i - y_m)$ precisely corresponds to $f'(y_i)$, we see that the equation $\sum_{i=1}^{d} A_{ji} (A^{-1})_{ik} = \mathbf{1}_{j=k}$ is really the same as the relation (5.12). This completes the proof. □

Remark 5.20. Note that one might worry whether the rational function Du can be evaluated at the point $(x, y_1, \dots, y_d)$, i.e. whether the denominators $\partial_Y f(x, y_i)$ vanish. However, since the roots $y_1, \dots, y_d$ of $f(x, Y)$ are distinct, the polynomial $f(x, Y)$ does not share zeros with its derivative $\partial_Y f(x, Y)$, showing that $\partial_Y f(x, y_i) \neq 0$.

Note that Proposition 5.19 only deals with first order derivatives. However, if we are to show that the vanishing multiplicities of $c_t(X,Y)$ carry over to $h_t(xX)$, we need a similar statement for higher derivatives. We will see in the following proposition that this is not much harder than the case we dealt with already.

Proposition 5.21. Let $x \in \mathfrak{S}$ and $y_1, \dots, y_d$ be the roots of $f(x,Y)$. Let u, v, h_t be defined through the relations (5.4), (5.5), (5.6) through the process of $f(X,Y)$-elimination on c_t. Then, for all $\ell \geq 0$ we have

$$E^{(\ell)}h_t(x) = E^{(\ell)}u(x, y_1, \dots, y_d).$$

Proof. To prove that the corresponding hyperderivates are equal, we will lift to the characteristic zero field and prove the equality there. Hence, let $\widehat{x}$ be a lift of x in K and let $\widehat{f}$ be the lift of f in $K[X,Y]$. Then, we define $\widehat{y}_1, \dots, \widehat{y}_d$ as the roots of $\widehat{f}(\widehat{x}, Y)$. Note that $\widehat{y}_1, \dots, \widehat{y}_d \in K$ must project to $y_1, \dots, y_d \in \mathbb{F}_q$ (up to rearrangement) and hence they must be different. Finally, let $\widehat{u}$ be a lift of u in $K[X, Y_1, \dots, Y_d]$ such that $\widehat{u}$ is symmetric in $Y_1, \dots, Y_d$. Then we obtain $\widehat{v}$ and $\widehat{h}_t$ from $\widehat{u}$ through the procedure analogous to the one described already. Note that we have $\pi(\widehat{v}) = v$ and $\pi(\widehat{h}_t) = h_t$.

Having made these definitions, the goal is now to prove $E^{(\ell)}\widehat{h}_t(\widehat{x}) = E^{(\ell)}\widehat{u}(\widehat{x}, \widehat{y}_1, \dots, \widehat{y}_d)$. Since we are working in characteristic zero, we can replace $E^{(\ell)}$ by $\frac{1}{\ell!}D^\ell$ and prove the statement by iterating Proposition 5.19.

Note that the proof of Proposition 5.19 did not rely on any assumptions regarding the base field. Hence, the case $\ell = 1$ follows directly from it. In other words, we have the relation $D\widehat{h}_t(\widehat{x}) = D\widehat{u}(\widehat{x}, \widehat{y}_1, \dots, \widehat{y}_d)$. Now, we will show how to iterate this to obtain the case $\ell = 2$ (and all higher ℓ will follow similarly). If we started with

$$\widehat{u}^{(1)}(X, Y_1, \dots, Y_d) = \sum_{i=1}^{d} D\widehat{c}(X, Y_i) \prod_{j \neq i} \widehat{c}(X, Y_j) = D\widehat{u}(X, Y_1, \dots, Y_d),$$

which is a symmetric polynomial in $Y_1, \dots, Y_d$ and constructed polynomials

$$\widehat{v}^{(1)}(X, \sigma_1(\boldsymbol{Y}), \dots, \sigma_d(\boldsymbol{Y})) = \widehat{u}^{(1)}(X, Y_1, \dots, Y_d) \qquad \widehat{h}^{(1)}(X) = \widehat{v}^{(1)}(X, -g_1(X), \dots, (-1)^d g_d(X)),$$

we would have $\widehat{h}^{(1)}(\widehat{x}) = D\widehat{h}_t(\widehat{x})$ for all $\widehat{x} \in \mathbb{Q}$. This means that we also have the equality between polynomials $\widehat{h}^{(1)}(X) = D\widehat{h}_t(X)$. Applying Proposition 5.19 to the polynomial $\widehat{h}^{(1)}$ implies that $D^2\widehat{h}_t(\widehat{x}) = D^2\widehat{u}(\widehat{x}, \widehat{y}_1 \dots, \widehat{y}_d)$, solving the $\ell = 2$ case. Repeating this procedure inductively now gives the proof for all ℓ. □

We have set the ground for showing the main property of h_t now - the high vanishing multiplicity.

Proposition 5.22. If h_t is a $f(X,Y)$-elimination of c_t and c_t vanishes to order M at all points (x,y) with $x \in \mathfrak{S}, y \in I_t(x)$, then h_t vanishes to order $M|I_t(x)|$ at x.

Proof. We need to check that $E^{(\ell)}h_t(x) = 0$ for all $0 \leq \ell < M|I_t(x)|$. Combining Proposition 5.21 and the product rule from Proposition 5.4, we find that

$$E^{(\ell)}h_t(x) = \sum_{\ell_1 + \dots + \ell_d = \ell} E^{(\ell_1)}c_t(x, y_1) \cdots E^{(\ell_d)}c_t(x, y_d).$$

If $\ell_1 + \dots + \ell_d = \ell < M|I_t(x)|$, the Pigeonhole principle implies there exists a term $E^{(\ell_i)}c_t(x, y_i)$ with $\ell_i < M$ and $y_i \in I_t(x)$. Since c_t vanishes to order M at (x, y_i), we find that $E^{(\ell_i)}c_t(x, y_i) = 0$. Hence, every term in the sum vanishes and we conclude $E^{(\ell)}h_t(x) = 0$, as needed. □

We are now ready for the proof of Proposition 5.13. Note that we have already derived Theorem 5.1 from Proposition 5.13 in Step 2. Therefore, with this proof also completes the proof of Theorem 5.1.

Proof of Proposition 5.13. By Proposition 5.22, we know that he auxiliary polynomial $h_t(X)$ vanishes to order $M|I_t(x)|$ at every $x \in \mathfrak{S}$, and therefore Corollary 5.7 shows that $\sum_{x \in \mathfrak{S}} M|I_t(x)| \leq \deg h_t$. Let us now bound the degree of h_t. We know $\deg u_t \leq d \deg c_t = d_t qM + 2qdd_t$. Furthermore, since $\deg g_i \leq i = \deg \sigma_i(\boldsymbol{Y})$, we find that $\deg h_t \leq \deg u_t \leq d_t qM + 2qdd_t$. Finally, since f does not divide c_t, we know that h_t is not identically zero.

Hence, we can conclude that $M \sum_{x \in \mathfrak{S}} |I_t(x)| \leq d_t qM + 2d^2 q$, and hence if we choose $M \geq \Omega(\frac{q^{1/2}}{d})$ we have

$$\sum_{x \in \mathfrak{S}} |I_t(x)| \leq d_t q + O(d^3 q^{1/2}).$$

□

Chapter 6

Bombieri's approach to Stepanov's method

In this section, we outline another approach to the proof of the Riemann Hypothesis for curves. This proof was initially suggested by Bombieri [4] and we follow its adaptation from the book of Niederreited and Xing [22]. In this chapter, unlike Chapter 5, we work with a smooth projective plane curve.

We will start by providing some geometric motivation, following Tao's blog post [32], and then we will present the formal proof phrased in terms of the function fields. Still, the main tool behind the proof is the polynomial method and therefore we will be able to bound the number of points on our curve from above. Unfortunately, using the polynomial method we will obtain only the upper bounds on this number, which will not be sufficient to prove the Riemann Hypothesis for curves. We will briefly indicate how one might transform these upper bounds into lower bounds, but this is be outside of the scope of our presentation.

We will work with a projective plane curve $C/\mathbb{F}_q$, where $q = s^2$ is a square and satisfies $q > (g+1)^4$. For the time being, the reader might imagine the case when C is a curve over $\mathbb{C}$ or $\mathbb{R}$ to obtain some intuition from the geometric picture. One important ingredient of our proof will be the notion of a *polynomial on* C.

Definition 6.1. Let P_∞ be a fixed rational point on $C/\mathbb{F}_q$. We say that a rational function $\varphi \in \mathbb{F}_q(C)$ is a *polynomial of degree* n if $\mathrm{div}_\infty(\varphi) = -nP_\infty$.

A way to think about this definition is setting P_∞ to be the "point at infinity". For example, if we consider the standard polynomials $p \in \mathbb{C}[x]$ and extend them to rational functions on the Riemann sphere $\mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$, the corresponding rational function has pole of order $\deg p$ at ∞.

From the definition, it is not hard to see that the set of polynomials of degree $\leq n$ on C, denoted by $\mathcal{P}_n$, is precisely the Riemann-Roch space $\mathcal{L}(nP_\infty)$. This means that we can compute the dimensions of the polynomial spaces accurately using the Riemann-Roch theorem.

Let us now present the setup of the proof. As before, we will characterize the $\mathbb{F}_q$-rational points of C using the Frobenius automorphism, i.e. we will try to find the number of points $P \in C$ satisfying $P^q = P$, where P^q denotes the point obtained by raising all coordinates of P to q-th power. Our idea will be to consider the surface $C \times C$ and two special curves on it, $C_1 = \{(P, P^s) : P \in C\}$ and $C_2 = \{(P^s, P) : P \in C\}$.[1] The motivation behind defining such C_1, C_2 is that the $\mathbb{F}_q$-rational points of $P \in C$ correspond to a point in the intersection $C_1 \cap C_2$. It is not hard to see this - if $P \in C$ and $P^q = P$, then $(P, P^s) \in C_1$ and since $P^s \in C$ we have $(P^q, P^s) = (P, P^s) \in C_2$. Thus, our goal will be to bound the size of the intersection $C_1 \cap C_2$ from above and this is what we use the polynomial method for.

We will be interested in finding a polynomial on $C \times C$ whose restriction to C_1 vanishes identically and whose restriction to C_2 does not vanish. Then, the size of the intersection $C_1 \cap C_2$ will be bounded by the degree of this polynomial.

The polynomials on $C \times C$ will be given by pairs (φ, ψ), where φ, ψ are polynomials on C and (φ, ψ)

[1] Technically, we are assuming that $P \mapsto P^s$ takes points of C to C again. This is true if the coefficients of the defining equation of C lie in $\mathbb{F}_s$, which is something we may assume.

evaluates to $\varphi(P_1)\psi(P_2)$ at a point $(P_1, P_2) \in C \times C$. In particular, we will be interested in the set $\mathcal{P}_\ell \otimes \mathcal{P}_n$ of polynomials (φ, ψ) on $C \times C$ where $\varphi \in \mathcal{P}_\ell, \psi \in \mathcal{P}_n$. When we consider the restrictions of such polynomial to C_1, we find that their degree is at most $\ell + sn$, while the restriction to C_2 will have degree at most $s\ell + n$. Now, if ℓ and n have different size, one might hope that the map from $\mathcal{P}_\ell \otimes \mathcal{P}_n$ to polynomials on C_2 might be injective, while the map to polynomials on C_1 might not be injective for dimension reasons. This would give a nonzero polynomial of degree $\leq s\ell + n$ on C_2 which vanishes on C_1, giving us a bound on the size of $C_1 \cap C_2$.

Of course, we need to solve some issues before formalizing this proof. Namely, we have developed a theory of projective plane curves, and it is not clear how whether one can think of C_1, C_2 as projective plane curves. Furthermore, one could ask whether the genus of C_1, C_2 equals the genus of C, which would be important if we wanted to apply the Riemann-Roch theorem to compute the dimension of the polynomial spaces. Although none of these issues are fundamental, developing the machinery from algebraic geometry that we need would simply took us to far from our topic. Fortunately, we can get around these issues by only considering the spaces of polynomials on C which imitate the construction of C_1, C_2, thus fitting this proof into the theory we have developed so far. Here is how we formally establish the upper bound, by working only within the function field.

Proposition 6.2. Let $C/\mathbb{F}_q$ be a smooth absolutely irreducible projective plane curve and assume that q is a square having $q > (g+1)^4$. Then, the number of $\mathbb{F}_q$-rational points of C is bounded by

$$\#C/\mathbb{F}_q \leq q + (2g+1)q^{1/2}.$$

Proof. The central object will be the space $\mathcal{P}_{s-1} \otimes \mathcal{P}_n$, where $n = s + 2g$ and $s = q^{1/2}$, which intuitively corresponds to polynomials on $C \times C$. Let us now define the spaces corresponding to the polynomials on C_1, C_2. We define the polynomial space V to be the image of the homomorphism $T : \mathcal{P}_{s-1} \otimes \mathcal{P}_n \to \mathcal{P}_{s(n+1)-1}$ given by $\varphi \otimes \psi \mapsto \varphi\psi^s$. More explicitly, V is simply the span of polynomials of the form $\varphi\psi^s$, where $\varphi \in \mathcal{P}_{s-1}, \psi \in \mathcal{P}_n$. It is not hard to see that if $\varphi \in \mathcal{P}_{s-1}, \psi \in \mathcal{P}_n$ then $\nu_{P_\infty}(\varphi\psi^s) = \nu_{P_\infty}(\varphi) + s\nu_{P_\infty}(\psi) \leq s - 1 + sn$, and so $\varphi\psi^s \in \mathcal{P}_{s(n+1)-1}$. The space V corresponds to the space of polynomials on C_1.[2]

Similarly, we define the space W as the image of the homomorphism $S : \mathcal{P}_{s-1} \otimes \mathcal{P}_n \to \mathcal{P}_{s(s-1)+n}$ given by $\varphi \otimes \psi \mapsto \varphi^s\psi$, which corresponds to the polynomials on C_2.

As explained in the motivation, our goal is to show that T is injective and that S not is injective. Let us begin by showing that T is injective. To do this, we choose the bases $\{\varphi_i\}$ and $\{\psi_j\}$ of $\mathcal{P}_{s-1}$ and $\mathcal{P}_n$ such that all their elements have distinct degrees. Then, suppose for contradiction that there is a nonzero element $\mu = \sum_{i,j} c_{i,j}\varphi_i \otimes \psi_j$ in the kernel of T, for some $c_{i,j} \in \mathbb{F}_q$. In other words, we have $\sum_{i,j} c_{i,j}\varphi_i\psi^s = \sum_i \varphi_i \left(\sum_j c_{i,j}^s \psi_j\right)^s = 0$, since s is a power of the characteristic of the field $\mathbb{F}_q$. However, note that the terms with different i in the sum are polynomials of different degrees. This is simply because modulo s the degree of $\varphi_i \left(\sum_j c_{i,j}\psi_j\right)^s$ is $\nu_P(\varphi_i)$. However, the strict triangle inequality (Proposition 2.9) for the valuation ν_P shows that nonzero elements of different valuations cannot sum to zero, which completes the proof of injectivity. Hence, the conclusion is that $T : \mathcal{P}_{s-1} \otimes \mathcal{P}_n \to V$ is a isomorphism.

Let us now show that $S : \mathcal{P}_{s-1} \otimes \mathcal{P}_n \to \mathcal{P}_{s(s-1)+n}$ is not injective using dimension counting. Since $\mathcal{P}_n = \mathcal{L}(nP_\infty)$, Riemann's inequality states that $n + 1 \geq \dim_{\mathbb{F}_q} \mathcal{P}_n \geq n + 1 - g$. Then, the following simple calculation shows that S must be injective,

$$\dim_{\mathbb{F}_q} \mathcal{P}_{s-1} \otimes \mathcal{P}_n \geq (s-g)(n+1-g) = (q^{1/2}-g)(q^{1/2}-g+1) = q + q^{1/2} - g(g+1) > q+g+1 = \dim_{\mathbb{F}_q} \mathcal{P}_{s^2-s+n}.$$

Note that the last equality follows since $\dim_{\mathbb{F}_q} \mathcal{P}_{s^2-s+n} = \dim_{\mathbb{F}_q} \mathcal{L}((s^2-s+n)P) = \dim_{\mathbb{F}_q} \mathcal{L}((q+2g)P) = q+g+1$. We conclude that S is not injective, an that its kernel contains an element $\mu \in \ker S$.

Now, we come to the final part of the argument. We claim that $\mu(Q) = 0$ for all $\mathbb{F}_q$-rational points of $C/\mathbb{F}_q$ except P_∞. To show this, write $\mu = \sum_i \varphi_i\psi_i^s$ for some $\psi_i \in \mathcal{P}_n$ (not necessarily the original elements of the basis), and note that μ must be regular at Q since $\operatorname{div}_\infty(\mu)$ is supported on P_∞. Since Q is $\mathbb{F}_q$-rational, we know that $\varphi_i(Q), \psi_i(Q) \in \mathbb{F}_q$ and therefore we have the following computation:

$$\mu(Q)^s = \left(\sum_i \varphi_i(Q)\psi_i(Q)^s\right)^s = \sum_i \varphi_i(Q)^s\psi_i(Q)^q = \sum_i \varphi_i(Q)^s\psi_i(Q) = 0,$$

[2] In the geometric motivation, we said that the polynomial corresponding to (φ, ψ) on C_1 should be given by $\varphi(P)\psi(P^s)$, but it is much easier for us to work with $\varphi(P)\psi(P)^s$. This is only a minor change and does not affect the course of the proof.

since $\mu \in \ker S$. The conclusion is that the number of $\mathbb{F}_q$-rational points of $C/\mathbb{F}_q$ is at most the number of zeros of μ, which can be bounded by its degree since $\deg \operatorname{div}_0(\mu) = \deg \operatorname{div}_\infty(\mu) \leq s - 1 + sn$. Therefore, we obtain $\#C/\mathbb{F}_q \leq 1 + s - 1 + s(s + 2g) \leq q + (2g + 1)s$. This completes the proof. □

Note that the bound from Proposition 6.2 is one-sided and bounds $\#C/\mathbb{F}_q$ only from above. Even when combined with the trace formula, this does not suffice to give us the full Riemann Hypothesis for curves. Hence, we need an additional argument to convert the upper bounds into lower bounds. Before that, let us briefly remind ourselves how we obtained lower bounds using Schmidt's approach. Along with this reminder, we will perform a brief comparison between the two methods.

One of the drawbacks of Schmidt's method is that it works very explicitly with the coordinates of the points on the curve $C/\mathbb{F}_q$, and that it treats the coordinates very asymmetrically. In particular, we count the number of solutions to $f(x, y) = 0$ when $x \in \mathbb{F}_q$ and y either belongs or does not belong to $\mathbb{F}_q$. The key to obtaining lower bounds is that we know that, without any constraint on y, the equation $f(x, y) = 0$ has $dq + O(d^2)$ solutions for $x \in \mathbb{F}_q$. Hence, by obtaining an upper bound on the number of solutions with $x \in \mathbb{F}_q, y \notin \mathbb{F}_q$, we automatically obtain a lower bound on the number of solutions $x \in \mathbb{F}_q, y \in \mathbb{F}_q$. Note that the key to transferring these bounds was that passing to $(x, y) \in \mathbb{F}_q \times \overline{\mathbb{F}_q}$ allowed us to count the number of solutions almost exactly.

However, Bombieri's approach works exclusively with the function field of the curve, without referencing the coordinates in the way Schmidt's approach does. This means that Bombieri's approach is intrinsic to the curve and that it treats the coordinates in a symmetric fashion. Hence, it is not obvious how to extend the curve C such that we are able to count the number of solutions exactly. To do this, one usually uses Galois coverings of curves, and this approach is described in [22], [4], [9]. Since discussing this approach in more detail would require the development of further machinery beyond the scope of this presentation, we do not choose to present it here.

Bibliography

[1] D. Allcock, Hilbert's Nullstellensatz, web.ma.utexas.edu/users/allcock/expos/nullstellensatz3.pdf, accessed on April 29, 2023.

[2] E. Artin, Quadratische Körper im Gebiete der höheren Kongruenzen, Jahrbuch der Philosophischen Fakultät zu Leipzig 1921.

[3] M. Atiyah, I. Macdonald, Introduction to Commutative Algebra, Reading: Addison-Wesley, 1969.

[4] E. Bombieri, Counting Points on Curves Over Finite Fields (d'après S. A. Stepanov). Séminaire Bourbaki vol. 1972/73 Exposés 418–435. Lecture Notes in Mathematics, vol 383. Springer, Berlin, Heidelberg.

[5] J. Bourgain, A. Gamburd, P. Sarnak, Markoff triples and strong approximation, C. R. Math. Acad. Sci. Paris 354 (2) (2016) 131–135.

[6] J. Bourgain, S. V. Konyagin, Estimates for the number of sums and products and for exponential sums over subgroups in fields of prime order, C. R. Math. Acad. Sci. Paris 337 (2003), 75–80.

[7] M, Baker, S. Norine, Riemann–Roch and Abel–Jacobi theory on a finite graph, Advances in Mathematics, vol. 215 (2007) 766-788.

[8] X. Chen, N. Kayal, A. Widgerson, Partial Derivatives in Arithmetic Complexity and Beyond. Foundations and Trends in Theoretical Computer Science, vol. 6 (2011) 1-138.

[9] P. Corvaja, U. Zanier, Greatest common divisors of $u-1$, $v-1$ in positive characteristic and rational points on curves over finite fields, J. Eur. Math. Soc. (JEMS) 15 (2013) 1927-1942.

[10] F. Enescu, Lecture notes on Commutative Algebra, https://math.gsu.edu/fenescu/commalglectures/Lect6.pdf, accessed on April 23, 2023.

[11] W. Fulton, Algebraic Curves, An Introduction to Algebraic Geometry. New York: Benjamin, 1969.

[12] M. van Frankehuijsen, The Riemann Hypothesis for function fields over a finite field. ArXiv preprint, arXiv:0806.0044, accessed on April 26, 2023.

[13] R. Griffon, Lecture notes for the course "Curves over Finite Fields", http://math.richardgriffon.me/CFF.html, accessed on March 15, 2023.

[14] R. Hartshorne, Algebraic Geometry. New York: Springer, 1977.

[15] H. Hasse, Beweis des Analogons der Riemannschen Vermutung für die Artinschen und F. K. Schmidtschen Kongruenzzetafunktionen in gewissen elliptischen Fällen. Vorläufige Mitteilung. Nachr. Ges. Wiss. Göttingen, Math.–Phys. Kl. I 1933(42), 253–262 (1933).

[16] D.R. Heath-Brown, S. Konyagin. New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum, Quarterly journal of mathematics Vol. 51, (2000), 221-235.

[17] Hilbert's Basis Theorem article on Art of Problem Solving Wiki. https://artofproblemsolving.com/wiki/index.php/Hilbert%27s_Basis_Theorem, accessed on April 23, 2023.

[18] D. Marcus, Number Fields. New York: Springer, 1977.

[19] E. Mazzoni, M. Schiltknecht, Cubic curves over finite fields, preprint, https://people.math.ethz.ch/ mschwagen/ellipticcurves2020/talk5finitefields.pdf, accessed on April 27, 2023.

[20] J. S. Milne, The Riemann Hypothesis over Finite Fields, from *The legacy of Bernhard Riemann after one hundred and fifty years*, edited by L. Ji, F. Oort, S.T. Yau. Somerville: International Press, 2015.

[21] D. Mumford, Abelian Varieties. London: Oxford University Press, 1970.

[22] H. Niederreiter and C. Xing, Algebraic Geometry In Coding Theory and Cryptography. Princeton: Princeton University Press, 2009.

[23] P. Roquette, The Riemann hypothesis in characteristic p in historical perspective. Springer, 2018.

[24] F. K. Schimdt, Analytische Zahlentheorie in Körpern der Charakteristik p, Mathematische Zeitschrift, vol. 33 (1931) 1-32.

[25] W. M. Schmidt, Equations Over Finite Fields: An Elementary Approach. Berlin: Springer-Verlag, 1976.

[26] I. R. Shafarevich, Basic Algebraic Geometry 1: Varieties In Projective Space. Berlin: Springer, 2013.

[27] J. Silverman, J. Tate. Rational Points On Elliptic Curves. New York: Springer, 1992.

[28] S. Stepanov, The number of points of a hyperelliptic curve over a finite prime field, (in Russian) Izv. Akad. Nauk SSSR Ser. Mat. 33 (1969), 1171–1181.

[29] Stichtenoth, H., Algebraic Function Fields and Codes, Berlin: Springer-Verlag, 1993.

[30] K.O. Stöhr, J. F. Voloch, Weierstrass Points and Curves Over Finite Fields. Proceedings of the London Mathematical Society Vol. 3s-52, (1986) 1-19.

[31] P. Swinnerton-Dyer, Applications of algebraic geometry to number theory, in *1969 Number Theory Institute*. Providence: American Mathematical Society, 1971.

[32] T. Tao, The Bombieri-Stepanov proof of the Hasse-Weil bound, blogpost terrytao.wordpress.com/2014/05/02/the-bombieri-stepanov-proof-of-the-hasse-weil-bound/, accessed on April 23, 2023.

[33] T. Tao, Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory, EMS Surveys in Mathematical Sciences (2014) 1-46.

[34] A. Weil, Sur les fonctions algebriques á corps de constantes fini. C. R. Acad. Sci. Paris 210, (1940). 592–594.

[35] A. Weil, On the Riemann hypothesis in function-fields. Proc. Nat. Acad. Sci. USA 27, (1941), 345–347.

[36] A. Weil, Sur les courbes algebriques et les variétés qui s'en déduisent. Actualités Sci. Ind., no. 1041 (1945).

[37] A. Weil, Varietés abéliennes et courbes algébriques. Actualités Sci. Ind., no. 1064, (1946).