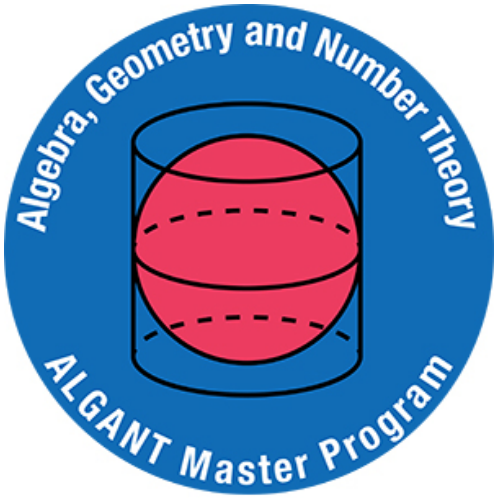
Exponential sums over finite fields and Stepanov's method

Diego Chicharro



Diego Chicharro

Master's Thesis

Exponential Sums over Finite Fields and Stepanov's Method

Supervised by Dr. Florent Jouve

2020 – 2021

université de BORDEAUX

UNIVERSITÉ FRANCO ITALIENNE

UNIVERSITÀ ITALO FRANCESE

Contents

Introduction

Exponential sums are certain arithmetically meaningful finite sums of roots of unity, which are omnipresent in distinct parts of number theory. One of the first historical examples, the Gauss sums, were introduced by C. F. Gauss to prove the quadratic reciprocity law differently from the other proofs that he had already given before, which were ingenious but not particularly illuminating, and they shed light upon the deep connections that this theorem has with cyclotomy. Furthermore, they proved to be an exceptionally useful tool in the study of similar problems, and they were fruitfully used to derive generalisations such as the cubic and biquadratic reciprocity laws and Eisenstein's reciprocity law, to mention only a few. Their usefulness is not restricted to this but is wide-ranging; for instance, they appear in the functional equations of the Dirichlet L-series, and they can be used to compute or estimate in an elementary way the number of solutions to some polynomial equations over finite fields.

Other interesting but more complicated exponential sums are the Kloosterman sums, which first appeared in the coefficients of certain modular forms in a paper by Poincaré, and can be considered as the discrete analogue of the Bessel function. They are named after Kloosterman, however, because he was the one that proved the first nontrivial bound for them, which he used to study the problem of representability of integers by quadratic forms in four variables, having solved the problem in five and more variables a few years earlier in his doctoral dissertation. Nevertheless, although the bound he found was enough for his purposes, it was far from optimal.

When one has to deal with an exponential sum, the ideal situation would be to be able to evaluate it explicitly, but this rarely occurs except in very particular cases. Fortunately, it is in general enough for applications to find good upper bounds, and different techniques have been developed to approach this problem. One of the most successful ones is based on how deeply connected exponential sums are to curves over finite fields. Indeed, finding an upper bound for an exponential sum is usually tantamount to estimating the number of rational points on a certain curve, and this latter problem can be studied with powerful tools from algebraic geometry.

More precisely, if X is an algebraic curve defined over a finite field of q elements, then the HASSE-WEIL bound says that the number of rational points on X differs from q by some quantity bounded by $2g\sqrt{q}$, where g is the genus of X. This celebrated result, first proved by WEIL, is equivalent to a version of the RIEMANN hypothesis for curves, which states that the zeros of a certain zeta function associated to the curve all lie on the line $\mathrm{Re}(s) = \frac{1}{2}$. Interestingly enough, a few years after WEIL's proof was published the Russian mathematician STEPANOV found a completely elementary way of proving the RIEMANN hypothesis by a simple counting argument. His method fundamentally consists in the following simple idea. If A is some finite subset of a field, say, and you can find a nonzero polynomial vanishing with order ℓ at all points of A, then the cardinality of A is trivially bounded by the degree of the polynomial over ℓ. It turns out that this bound is strong enough to deduce the required estimates.

While STEPANOV only applied his method to a certain family of hyperelliptic curves, it was afterwards generalised by BOMBIERI to arbitrary curves over finite fields, at the expense of using the RIEMANN-ROCH theorem. Understanding BOMBIERI's proof is the aim of the first chapter, which begins with a somewhat lengthy discussion on the theory of curves over finite fields and a sketch of the proof of the RIEMANN-ROCH theorem. We then define the zeta function associated to an algebraic curve, prove both the rationality and the functional equation it satisfies, and finally show how STEPANOV's method is used to prove the RIEMANN hypothesis.

In the second chapter we deal with two important families of exponential sums, namely KLOOSTERMAN sums and WEIL sums. Their study parallels that of curves in the first chapter in the sense that one constructs a zeta function associated to them which satisfies the expected properties, namely that it is a rational function, it has some functional equation, and that the RIEMANN hypothesis holds. However, in contrast to the case of curves, where the RIEMANN-ROCH theorem plays a leading role, this chapter is computational in nature and everything is explicit. Indeed, we prove again the RIEMANN hypothesis for the family of curves that arise in this context following essentially STEPANOV's original method. The proof crucially depends upon finding an auxiliary polynomial vanishing at the points on the curve with large order, and the construction is largely inspired by diophantine approximation: one considers a polynomial of large degree with indeterminate coefficients, and then imposes the vanishing of all the derivatives of the polynomial at these points up to certain fixed order. A subtlety to be considered here is that we are working in a field of positive characteristic, so ordinary

derivatives no longer can be used to determine the order of a zero. Nevertheless, the so-called HASSE derivatives solve this problem. Finally taking some adequate values of the parameters one can ensure that there are more variables than (linear) conditions, so a nontrivial solution exists by linear algebra. As an application of the RIEMANN hypothesis we find the optimal bound for KLOOSTERMAN sums, which is the most important theorem of the chapter.

An important point to be taken into account is that even though the HASSE-WEIL bound is applicable in a wide variety of situations, it becomes useless if the genus of the curve is large compared with the cardinality of the finite field. This happens, for example, for the so-called HEILBRONN sums, which we study in the third chapter. The algebraic curves associated to these sums have genera that grow quickly with the order of the field, and as a consequence one obtains worse-than-trivial estimates. It was not until 1996 that HEATH-BROWN, applying in a striking way a modified version of STEPANOV's method, found the first nontrivial upper bound for these sums. The key ingredient was the estimation of the number of solutions to the equation $f(X) = a$ over a finite field of p elements, where f is a truncation of the power series expansion of the logarithm. Four years later he and KONYAGIN refined the method to obtain a better bound, and SHKREDOV subsequently made further improvements, although we are still far from the optimal bound.

Up to this point we have talked about exponential sums and their connection to algebraic curves, but there is a natural generalisation that arises in different contexts, namely exponential sums in several variables. These sums are linked to higher dimensional algebraic varieties over finite fields, and as in the one-dimensional case one can construct a zeta function associated to a variety and formulate analogous statements about it, which are known as the WEIL conjectures. The rationality and the functional equation having being proved by DWORK and GROTHENDIECK in the sixties, it was not until 1974 that DELIGNE proved the RIEMANN hypothesis using deep results from algebraic geometry. In the fourth and last chapter we discuss, although only briefly, DELIGNE's theory formulated for exponential sums, and we examine an interesting example of a three-dimensional exponential sum studied by BIRCH and BOMBIERI, which exemplifies how powerful DELIGNE's results are.

Acknowledgments. I would like to thank my advisor Dr. Florent Jouve for suggesting a very interesting topic for the master's thesis, for his careful reading of the manuscript, and for his invaluable suggestions to improve it.

CHAPTER 1

Curves over finite fields

1.1. Algebraic curves

1.1.1. Basic definitions. Let k be a field and X be a scheme over k. The structure sheaf of X is denoted by $\mathcal{O}_X$, and for a point $x \in X$ we let:

$\mathcal{O}_{X,x}$ be the local ring at x,
$\mathfrak{m}_x$ be the unique maximal ideal of $\mathcal{O}_{X,x}$, and
$k(x) = \mathcal{O}_{X,x}/\mathfrak{m}_x$ be the residue field at x.

Recall that X is *normal* if $\mathcal{O}_{X,x}$ is a normal domain for all $x \in X$, i.e., it is integrally closed in its fraction field. This condition is equivalent to the normality of the ring $\mathcal{O}_X(U)$ for all affine subsets $U \subseteq X$. We say that X is *projective* if it is a closed subscheme of some projective space $\mathbb{P}^n_k = \operatorname{Proj} k[T_0, \ldots, T_n]$, in which case there exists some homogeneous ideal I of $k[T_0, \ldots, T_n]$ such that $X = \operatorname{Proj} k[T_0, \ldots, T_n]/I$. A closed point $x \in X$ is said to be *nonsingular* if the local ring $\mathcal{O}_{X,x}$ is regular, this is, it is a Noetherian local ring such that $\dim_k \mathfrak{m}_x/\mathfrak{m}_x^2$ coincides with the Krull dimension of $\mathcal{O}_{X,x}$. When $\dim \mathcal{O}_{X,x} = 1$ this last condition is equivalent to saying that $\mathcal{O}_{X,x}$ is normal, or that it is a discrete valuation ring. We say that X is *regular* or *nonsingular* if every point has an affine open neighbourhood $U \subseteq X$ such that $\mathcal{O}_X(U)$ is regular. Equivalently, X is locally Noetherian and nonsingular at every closed point. If X is integral, we define its *function field* $k(X)$ as the residue field of the generic point of X. If $U \subseteq X$ is an affine open subset, then the natural map $\mathcal{O}_X(U) \to k(X)$ is injective and induces an isomorphism between the fraction field of $\mathcal{O}_X(U)$ and the function field of X.

1.1.2. Base change. Let X be a scheme over a field k. If K/k is a field extension we denote by X_K the base change $X \times_k \operatorname{Spec} K$. Recall that by the universal property of the fibre product we have a bijection $X_K(K) = X(K)$ between K-rational points of X_K and K-rational points of X. Furthermore, k-rational points of X are naturally identified with points $x \in X$ with residue field isomorphic to k. More generally, there is a one-to-one correspondence between K-rational points of X and pairs (x, ι) of closed

points $x \in X$ and k-embeddings $\iota : k(x) \to K$. When X is of finite type over k we also know that a point $x \in X$ is closed if and only if the residue field $k(x)$ is a finite extension of k. The degree $[k(x) : k]$ is called the *degree* of x, and is denoted by $\deg x$.

Notice that the group $\mathrm{Aut}_k(K)$ of k-automorphisms of K acts on $X(K)$ in a natural way: a k-automorphism $\sigma : K \to K$ induces a morphism $\operatorname{Spec}\sigma : \operatorname{Spec} K \to \operatorname{Spec} K$ of k-schemes, and by composition we have the map $X(K) \to X(K), s \mapsto s \circ \operatorname{Spec}\sigma$. In particular, suppose that k is perfect and fix an algebraic closure $\overline{k}$ of k. Then $\mathrm{Gal}(\overline{k}/k)$ acts on $X(\overline{k})$, and the $\mathrm{Gal}(\overline{k}/k)$-orbits of $X(\overline{k})$ are exactly the closed points of X.

PROPOSITION 1.1.1. *There is a one-to-one correspondence between* $\mathrm{Gal}(\overline{k}/k)$*-orbits in* $X(\overline{k})$ *and closed points of* X *given by*

$$(s : \operatorname{Spec}\overline{k} \to X) \mapsto s(\operatorname{Spec}\overline{k}).$$

Furthermore, the degree of a closed point $x \in X$ *is exactly the cardinality of the corresponding orbit.*

PROOF. If we identify $X(\overline{k})$ with the set of pairs (x, ι) with $x \in X$ closed point and $\iota \in \mathrm{Hom}_k(k(x), \overline{k})$, then the $\mathrm{Gal}(\overline{k}/k)$-action is given by $\sigma(x, \iota) = (x, \sigma \circ \iota)$, and the claim follows as $\mathrm{Gal}(\overline{k}/k)$ acts transitively on $\mathrm{Hom}_k(k(x), \overline{k})$, which has cardinality $[k(x) : k]$ since $\overline{k}$ is algebraically closed. □

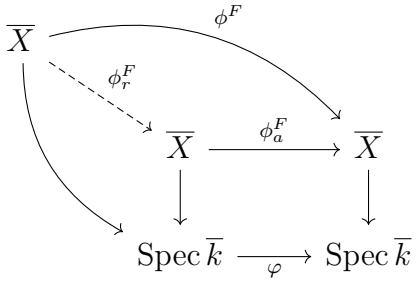
Thus for each closed point $x \in X$ there are $\deg x$ points on $X_{\overline{k}}$ lying above x. In addition, since the $\mathrm{Gal}(\overline{k}/k)$-orbit of $(x, \iota) \in X(\overline{k})$ is a singleton if and only if $\deg x = 1$, we see that $X(k) = X(\overline{k})^{\mathrm{Gal}(\overline{k}/k)}$. More generally, if K/k is an algebraic extension contained in $\overline{k}$, then K-rational points are in one-to-one correspondence with $\overline{k}$-rational points fixed by $\mathrm{Gal}(\overline{k}/K)$.

1.1.3. Frobenius morphisms. Let $k = \mathbb{F}_q$ be a field of q elements, fix an algebraic closure $\overline{k}$ of k, and let $\varphi : \operatorname{Spec}\overline{k} \to \operatorname{Spec}\overline{k}$ be the morphism induced by the FROBENIUS morphism $\overline{k} \to \overline{k}, \alpha \mapsto \alpha^q$. Also let X be a scheme over k and $\overline{X}$ be the base change $X \times_k \operatorname{Spec}\overline{k}$. We define the following endomorphisms of $\overline{X}$:

(1) The *absolute* FROBENIUS is the morphism ϕ^F that is the identity on points of $\overline{X}$ and that induces the map $f \mapsto f^q$ in $\mathcal{O}_{\overline{X}}(U)$ for all open subsets $U \subseteq \overline{X}$. It is not a morphism of $\overline{k}$-schemes.
(2) The *relative* FROBENIUS is the morphism $\phi_r^F = \phi^F \times \mathrm{id}_{\operatorname{Spec}\overline{k}}$ of $\overline{k}$-schemes.
(3) The *arithmetic* FROBENIUS is the morphism $\phi_a^F = \mathrm{id}_X \times \varphi$, which is an automorphism of X-schemes (but not of $\overline{k}$-schemes).

(4) The *geometric* FROBENIUS is the inverse of ϕ_a^F, this is, $\phi_g^F = \mathrm{id}_X \times \varphi^{-1}$.

They are related through the equalities $\phi^F = \phi_r^F \circ \phi_a^F = \phi_a^F \circ \phi_r^F$, so we have the commutative diagram



which in particular shows that the action of the morphism φ on closed points of $\overline{X}$ coincides with the action of the arithmetic FROBENIUS. Hence k-rational points on X correspond to closed points on $\overline{X}$ fixed by ϕ_a^F,

$$X(k) = \{x \in \overline{X}(\overline{k}) : \phi_a^F(x) = x\}.$$

These endomorphisms are functorial in the sense that for all morphisms $\phi : X \to Y$ of schemes, the diagram

$$\begin{array}{ccc} \overline{X} & \xrightarrow{\phi^F} & \overline{X} \\ {\scriptstyle \phi\times \mathrm{id}_{\overline{k}}}\downarrow & & \downarrow{\scriptstyle \phi\times \mathrm{id}_{\overline{k}}} \\ \overline{Y} & \xrightarrow[\phi^F]{} & \overline{Y} \end{array}$$

commutes, and similarly for ϕ_r^F, ϕ_a^F and ϕ_g^F.

1.1.4. Algebraic curves and function fields. The main objects we will work with in this Chapter are projective nonsingular (or, equivalently, normal) curves over a perfect field k, which is usually a finite field $\mathbb{F}_q$ or its algebraic closure $\overline{\mathbb{F}}_q$.

DEFINITION 1.1.2. A *curve* is a separated integral scheme of finite type over k of dimension one.

One can prove that there is a contravariant equivalence of categories between the category of projective nonsingular curves over k with nonzero morphisms of k-schemes (equivalently, finite morphisms, or surjective morphisms) and the category of function fields of transcendence degree one over k with k-algebra homomorphisms, see for example [27, Theorem 0BY1]. In particular, since every function field is a finite extension of the purely transcendental extension $k(t)$, which is the function field of $\mathbb{P}^1_k = \operatorname{Proj} k[T_0, T_1]$, every curve X admits a surjective morphism $X \to \mathbb{P}^1_k$. We also have that there is a one-to-one correspondence between closed points on X and discrete

valuation subrings of $k(X)$. Indeed, given a closed point x, the subring $\mathcal{O}_{X,x} \subseteq k(X)$ is a discrete valuation ring by definition of regularity. Conversely, we prove in Lemma 1.1.8 below that one can find a unique closed points x from a given discrete valuation subring R of $k(X)$ such that $\mathcal{O}_{X,x} = R$. The normalised valuation on $\mathcal{O}_{X,x}$ is the map $\mathrm{ord}_x : \mathcal{O}_{X,x} \to \mathbb{Z} \cup \{\infty\}$ sending an element $f \in \mathcal{O}_{X,x}$ to the supremum of the integers d such that $f \in \mathfrak{m}_x^d$. In the case of $X = \mathbb{P}^1_k$ we can view nonzero elements of $k(\mathbb{P}^1_k)$ as quotients of the form $f = g/h$ with $h, g \in k[T_0, T_1] \setminus \{0\}$ homogeneous of the same degree. Thus, if we factor them as product of irreducible polynomials, which correspond to discrete valuation subrings of $k(\mathbb{P}^1_k)$ and hence to closed points of $\mathbb{P}^1_k$, the fact that $\deg g = \deg h$ implies that

$$\sum_{x \text{ closed point}} \mathrm{ord}_x(f) \deg x = 0, \tag{1.1.3}$$

this is, f has as many zeros as poles, counted with the appropriate multiplicities. Indeed, one only needs to notice that the degree of a point is exactly the degree of the corresponding irreducible polynomial. Equation (1.1.3), which holds for any curve X, is known as the *product formula*.

PROPOSITION 1.1.4. *Let X be an algebraic curve over k and $f \in k(X)$. Then*

$$\sum_x \mathrm{ord}_x(f) \deg x = 0,$$

where x runs over all the closed points of X.

To prove Proposition 1.1.4 we recall first some definitions. Let $U \subseteq X$ be an affine open subset of X, and let $R = \mathcal{O}_X(U)$ be the ring of regular functions on U. As we remarked at the beginning of the section, the fraction field of R is naturally identified with $k(X)$, and furthermore R is a DEDEKIND domain. Indeed, it has KRULL dimension one since X is has dimension one, it is Noetherian since X is of finite type over k, and it is normal as X is nonsingular. In particular ideals in R factor in a unique way as product of prime ideals. If t is a transcendental element in R and $X \to \mathbb{P}^1_k$ is the morphism corresponding to $k(t) \to k(X)$, then $R/k[t]$ is a finite extension of DEDEKIND domains and we see that for a closed point $x \in X$ the numbers $\deg x$ and $\mathrm{ord}_x(t)$ correspond to the usual notions of inertia degree and ramification index respectively. More generally, let $\phi : X \to Y$ be a morphism of curves, which induces a k-algebra homomorphism $\phi^* : k(Y) \to k(X)$ between the function fields, and therefore a finite extension of fields $k(X)/\phi^* k(Y)$.

DEFINITION 1.1.5. The *degree* of $\phi : X \to Y$ is defined as $\deg \phi = [k(X) : \phi^* k(Y)]$.

In particular ϕ is an isomorphism if and only if $\deg \phi = 1$.

DEFINITION 1.1.6. Let $\phi : X \to Y$ be a morphism, $y \in Y$ be a closed point, and t_y be a *uniformiser* at y, this is, an element $t_y \in k(Y)$ with $\mathrm{ord}_y(t_y) = 1$. We say that y is *unramified* in X if $\mathrm{ord}_x(\phi^* t_y) = 1$ for all $x \in \phi^{-1}(y)$. Otherwise we say that y is *ramified*.

PROOF OF PROPOSITION 1.1.4. A well-known theorem on DEDEKIND domains says that, for a fixed closed point $y \in Y$,

$$\sum_{x \in \phi^{-1}(y)} \mathrm{ord}_x(\phi^* t_y)[k(x) : \phi^* k(y)] = \deg \phi, \tag{1.1.7}$$

where t_y is a fixed uniformiser at y. Notice that, since ϕ is a finite morphism, it has finite fibres and therefore this sum is well-defined. It is now immediate that, if $f \in k(X)$ and Y is the curve with function field $k(t, f)$, then

$$\sum_x \mathrm{ord}_x(f) \deg x = \sum_y \sum_{x \in \phi^{-1}(y)} \mathrm{ord}_x(f) \deg x = \deg \phi \sum_y \mathrm{ord}_y(f) \deg y$$

on using the formulas

$$\deg x = [k(x) : \phi^* k(y)] \deg y,$$
$$\mathrm{ord}_x(\phi^* f) = \mathrm{ord}_x(\phi^* t_y) \, \mathrm{ord}_y(f) \quad \text{for } f \in k(Y).$$

An important particular case of the latter equation is that $\mathrm{ord}_x(\phi^* f) = \mathrm{ord}_y(f)$ when ϕ is an isomorphism. Thus, to prove Proposition 1.1.4 it suffices to consider the case when $k(X) = k(t, f)$. If f is algebraic over k then it is integral over k, so it is contained in all discrete valuation rings $S \subseteq k(t, f)$. But the same is true for f^{-1}, so f is a unit in S. This means that $\mathrm{ord}_x(f) = 0$ for all x, and the product formula is trivially true for f. If f is not algebraic then $k(f)$ is a purely transcendental extension, so it is the function field of $\mathbb{P}^1_k$. Let $\phi_f : X \to \mathbb{P}^1_k$ be the corresponding morphism of curves and let R be the integral closure of $k[f]$ in $k(t, f)$. If y is the closed point of $\mathbb{P}^1_k$ corresponding to the prime ideal $fk[f]$ of $k[f]$, then by Equation (1.1.7) we have

$$\sum_{x \in \phi_f^{-1}(y)} \mathrm{ord}_x(f) \deg x = \deg \phi_f = [k(t, f) : k(f)].$$

Notice that if x is a closed point such that $\mathrm{ord}_x(f) \geq 0$, and it corresponds to some discrete valuation ring $S \subseteq k(t, f)$, then $f \in S$ and consequently $R \subseteq S$ since S

is integrally closed. Thus S corresponds to some prime of R lying above $fk[f]$, so $\mathrm{ord}_x(f) > 0$ exactly when $x \in \phi_f^{-1}(y)$. Similarly, arguing with f^{-1} we have

$$\sum_{x \in \phi_{1/f}^{-1}(z)} \mathrm{ord}_x(f^{-1}) \deg x = \deg \phi_{1/f} = [k(t,f) : k(f)]$$

where z is the closed point corresponding to $f^{-1}k[f^{-1}]$. Since $\mathrm{ord}_x(f)$ is nonzero if and only if $x \in \phi_f^{-1}(y)$ or $x \in \phi_{1/f}^{-1}(z)$, we deduce that

$$\sum_x \mathrm{ord}_x(f) \deg x = \sum_{x \in \phi_f^{-1}(y)} \mathrm{ord}_x(f) \deg x - \sum_{x \in \phi_{1/f}^{-1}(z)} \mathrm{ord}_x(f^{-1}) \deg x = 0.$$

This ends the proof of the proposition. □

LEMMA 1.1.8. *There is a one-to-one correspondence between closed points on X and discrete valuation subrings of $k(X)$.*

PROOF. We only need to find, for each discrete valuation subring $R \subseteq k(X)$, a closed point $x \in X$ with $\mathcal{O}_{X,x} = R$, and then show that distinct points have distinct local rings. Let f be a uniformiser for R and let $X \to \mathbb{P}^1_k$ be the morphism associated to the inclusion $k(f) \to k(X)$. Let $y \in Y$ be the prime corresponding to the maximal ideal $fk[f]$ of $k[f]$, and let S be the integral closure of $k[f]$ in $k(X)$. Since $f \in R$ and R is integrally closed it is clear that $S \subseteq R$. Now, the ideal $\mathfrak{m} = fR \cap S$ is maximal in S so it corresponds to some point x lying above y, and we claim that the local ring $\mathcal{O}_{X,x} = S_{\mathfrak{m}}$ is R. But $\mathcal{O}_{X,x}$ is a discrete valuation ring, and it is well-known that if $R_1 \subseteq R_2$ are discrete valuation rings, then we must have $R_1 = R_2$, see Chapter 5, p. 72 in [1]. Finally, suppose that $\mathcal{O}_{X,x} = \mathcal{O}_{X,x'}$, and let $\operatorname{Spec} A$ be an affine neighbourhood containing x and x'. Then $\mathfrak{m}_x = \mathfrak{m}_{x'}$ in A, so $x = x'$, as wanted. □

1.1.5. Galois covers. Let $\phi : X \to Y$ be a morphism of nonsingular projective curves over a field k, which endows X with a structure of Y-scheme. Then each $\phi^*k(Y)$-algebra automorphism $\sigma^{-1} \in \mathrm{Aut}_{\phi^*k(Y)}(k(X))$ of $k(X)$ induces a morphism $\phi_\sigma : X \to X$ of Y-schemes such that $\phi_\sigma^* = \sigma^{-1}$, and consequently $\mathrm{Aut}_{\phi^*k(Y)}(k(X))$ acts on the fibre of each closed point y of Y permuting the points lying above. The reason we use σ^{-1} instead of σ is that in this way we have $\sigma(\mathcal{O}_{X,x}) = \mathcal{O}_{X,\phi_\sigma(x)}$ for each $x \in X$.

DEFINITION 1.1.9. If the extension $k(X)/\phi^*k(Y)$ is GALOIS we say that $\phi : X \to Y$ is a GALOIS cover.

An important property of a GALOIS cover is that the GALOIS-action on the curve X is transitive on the points lying above a fixed $y \in Y$.

PROPOSITION 1.1.10. *Suppose $\phi : X \to Y$ is a* GALOIS *cover and let $y \in Y$. Then the action of $\mathrm{Gal}(k(X)/\phi^* k(Y))$ on $\phi^{-1}(y)$ is transitive, this is, for all pairs $x, x' \in \phi^{-1}(y)$ there exists some automorphism σ such that $\sigma(x) = x'$.*

PROOF. Let $\operatorname{Spec} A \subseteq Y$ be an affine neighbourhood of y, and let $\operatorname{Spec} B = \phi^{-1}(\operatorname{Spec} A)$. Also let $\mathfrak{p} \in \operatorname{Spec} A$ be the prime corresponding to y. If the action is not transitive then there are primes $\mathfrak{q}_1, \mathfrak{q}_2 \in \operatorname{Spec} B$ above $\mathfrak{p}$ such that $\sigma(\mathfrak{q}_1) \neq \mathfrak{q}_2$ for all automorphisms σ. Using the Chinese remainder theorem find some $\alpha \in \mathfrak{q}_2$ such that $\alpha \equiv 1 \pmod{\sigma^{-1}(\mathfrak{q}_1)}$ for all σ. Then $\beta = \prod_\sigma \sigma(\alpha) \in A$ is not in $A \cap \mathfrak{q}_1 = \mathfrak{p}$ since $\beta \equiv 1 \pmod{\mathfrak{q}_1}$. But $\beta \in A \cap \mathfrak{q}_2 = \mathfrak{p}$, which is a contradiction. □

PROPOSITION 1.1.11. *Suppose $\phi : X \to Y$ is a* GALOIS *cover and let $y \in Y$. Then the ramification index $\operatorname{ord}_x(\phi^* t_y)$ and inertia degree $[k(x) : \phi^* k(y)]$ do not depend on $x \in \phi^{-1}(y)$, they depend only on y.*

PROOF. Each $\phi^* k(Y)$-algebra automorphism $\sigma^{-1} : k(X) \to k(X)$ induces a $\phi^* k(y)$-algebra isomorphism $k(x) \to k(\phi_\sigma(x))$, so the inertia degrees of x and $\phi_\sigma(x)$ are equal. On the other hand, since $\phi^* = \sigma^{-1}\phi^*$ and ϕ_σ is an isomorphism,

$$\operatorname{ord}_x(\phi^* t_y) = \operatorname{ord}_x(\sigma^{-1}\phi^* t_y) = \operatorname{ord}_{\phi_\sigma(x)}(\phi^* t_y),$$

so the claim follows from the transitivity of the the action. □

1.1.6. Geometric properties of curves. Some important properties a scheme may have, such as irreducibility or integrality, are not preserved under base change in general. This motivates the following definition.

DEFINITION 1.1.12. A scheme X over a field k is called *geometrically reduced* (resp. *irreducible*, *integral*) over k if $X \times_k \operatorname{Spec} k'$ is reduced (resp. irreducible, integral) for all extensions k'/k.

One has that X is geometrically irreducible if and only if $X \times_k \operatorname{Spec} k'$ is irreducible for all finite separable extensions k'/k, see [27, Lemma 038I]. On the other hand, over perfect fields the condition of geometric reducibility is automatic.

LEMMA 1.1.13. *If X is a reduced scheme over a perfect field k, then X is geometrically reduced over k.*

PROOF. For simplicity we only prove it when X is integral and k'/k is an algebraic extension. The general case follows from Theorem 3, Chapter V, Section 15 in [5].

Let $\operatorname{Spec} A \subseteq X$ be an affine open subset. Then A is an integral domain, so it is a subring of its ring of fractions $K = k(X)$. Since any separable extension is direct limit of finite separable extensions and the direct limit of reduced rings is again reduced, it suffices to consider the case when k' is a finite separable extension. Write $k' = k(\alpha)$ for some separable element α. Tensoring with k' we find that $A \otimes_k k'$ is a subring of $K \otimes_k k' = K[T]/(f)$, where $f \in k[T]$ is the minimal polynomial of α. Since f is separable the ring $K[T]/(f)$ is reduced, and hence so is $A \otimes_k k'$. This proves that $X \times_k \operatorname{Spec} k'$ is reduced. □

Note however that if X is irreducible it might not be true that $X_{k'}$ is irreducible for some extension k'/k, even if k is perfect. For this reason when working with a curve X we shall also require the additional hypothesis that X is geometrically irreducible to guarantee that base extensions of X are also curves. This condition is equivalent to requiring that k is algebraically closed in the function field $k(X)$ of X.

PROPOSITION 1.1.14. *A curve X/k is geometrically irreducible if and only if k is algebraically closed in the function field $k(X)$.*

In the proof of the proposition we will use the following lemma.

LEMMA 1.1.15. *Let X be an integral scheme over a perfect field k. Then X is geometrically integral if and only if $\operatorname{Spec}(k(X)\otimes_k k')$ is irreducible for all finite separable extensions k'/k.*

PROOF. Let k'/k be a separable extension and suppose that $X' = X \times_k \operatorname{Spec} k'$ is integral. Then, since X' is covered by the affine schemes of the form $\operatorname{Spec}(A \otimes_k k')$ with $\operatorname{Spec} A \subseteq X$, we have

$$k(X') = \varinjlim_{\operatorname{Spec} A \subseteq X} A \otimes_k k' = k(X) \otimes_k k'$$

on using that the tensor product commutes with colimits. Thus $k(X) \otimes_k k'$ is a field, so $\operatorname{Spec}(k(X)\otimes_k k')$ is irreducible. Conversely, suppose that $\operatorname{Spec}(k(X)\otimes_k k')$ is irreducible. Since k' is separable this means that $k(X) \otimes_k k'$ is a field. The natural morphism $\operatorname{Spec} k(X) \to X$ induces a morphism $\operatorname{Spec}(k(X) \otimes_k k') \to X \times_k \operatorname{Spec} k'$ which factors through any affine subset $\operatorname{Spec}(A \otimes_k k')$ with $\operatorname{Spec} A \subseteq X$. Let ξ' be the unique point in the image of this morphism. Since A is a subring of $k(X)$, $A \otimes_k k'$ is a subring of $k(X) \otimes_k k'$, so ξ' corresponds to the zero ideal of $A \otimes_k k'$. Hence $\operatorname{Spec}(A \otimes_k k')$ is irreducible. Choosing now an affine cover $\{\operatorname{Spec} A_i\}_{i\in I}$ of irreducible subsets of X such

that $\operatorname{Spec} A_i \cap \operatorname{Spec} A_j$ is nonempty for all $i, j \in I$ and using the flatness of k' it follows that X' is also irreducible. □

Thus Proposition 1.1.14 will follow if we prove that $\operatorname{Spec}(k(X) \otimes_k k')$ is irreducible for all finite separable extensions k'/k, and this in turn is a particular case of the following lemma, where the field need not be perfect.

LEMMA 1.1.16. *Let k be any field and K/k be an extension. Then $\operatorname{Spec}(K \otimes_k k')$ is irreducible for all finite separable k'/k if and only if K is separably closed in k.*

PROOF. Suppose first that k is separably closed in K, and let $\alpha \in K$ be separable over k. If $f \in k[T]$ is the minimal polynomial of α over k, then

$$\operatorname{Spec}(K \otimes_k k(\alpha)) = \operatorname{Spec}(K \otimes_k (k[T]/(f))) = \operatorname{Spec}(K[T]/(f))$$

is not irreducible unless $\alpha \in k$ since $T - \alpha$ divides f in $K[T]$. Conversely, suppose that k is algebraically closed in K, and let k'/k be a finite separable extension. Then $k' = k(\alpha)$ for some α separable over k and $\operatorname{Spec}(K \otimes_k k') = \operatorname{Spec}(K[T]/(f))$, where f is the minimal polynomial of α, so we see that $\operatorname{Spec}(K \otimes_k k')$ is reducible if and only if f has a nontrivial factorisation $f = f_1 f_2$ in $K[T]$. But the coefficients of the polynomials f_1 and f_2 are separable over k since they lie in the extension of k generated by the roots of f. Thus, if k^{sep} is a fixed separable closure of k containing these coefficients, we have that $f_1, f_2 \in K[T] \cap k^{\text{sep}}[T] = k[T]$, contradicting that f is irreducible in $k[T]$. □

Finally, while the properties of being projective, separated, and of finite type are preserved under base change, it is not true in general for normality (nor for regularity, but this property is more subtle since local Noetherianity is not preserved under arbitrary base change to extensions k'/k either). Thus one can define similarly the notion of *geometric normality* of a scheme, but by Lemma 1.1.17 it is not interesting if the field k is perfect. Hence, since an application of NOETHER's normalisation lemma shows that the dimension is preserved under base change to k' (see Lemma 1.1.18 below), we have that the base change of a geometrically irreducible nonsingular curve over a perfect field k to an extension k'/k is again a geometrically irreducible nonsingular curves over k'.

LEMMA 1.1.17. *Let k be a perfect field and R be a k-algebra. Then R is normal if and only if $R \otimes_k k'$ is normal for all extensions k'/k.*

PROOF. We work with the additional hypothesis that k' is algebraic over k and that R is geometrically irreducible, which is the case we need. The proof of the general

statement can be found in [27, Lemma 037Z]. Writing k' as a direct limit of finite separable extensions of k and noting that normality is preserved under direct limits we only need to consider finite separable extensions k'/k. But if $k' = k[T]/(f)$ for some irreducible monic polynomial $f \in k[T]$, then $R \otimes_k k' = R[T]/(f)$, and f is still irreducible in $R[T]$ since R is geometrically irreducible. Hence, if α is a root of f, we have $R \otimes_k k' \simeq R[\alpha]$, and $R[\alpha]$ is integral in $\mathrm{Frac}(R)(\alpha) = \mathrm{Frac}(R[\alpha])$ since α is integral over R. □

LEMMA 1.1.18. *Let X be an integral scheme of finite type over k and let k'/k be an extension of fields. If $X_{k'}$ is irreducible then* $\dim X_{k'} = \dim X$.

PROOF. If $\{\mathrm{Spec}\, A_i\}_{i\in I}$ is an affine cover of X, $\{\mathrm{Spec}(A_i \otimes_k k')\}_{i\in I}$ is an affine cover of X' and $\dim X'$ is the supremum of the dimensions of $\mathrm{Spec}(A_i \otimes_k k')$. Thus we only need to show that $\dim(A_i \otimes_k k') = \dim A_i$. By NOETHER's normalisation lemma there is a finite injective morphism $k[T_1, \ldots, T_d] \to A_i$ of k-algebras, where $d = \dim A_i$. Since k' is a flat k-algebra we have that $k'[T_1, \ldots, T_d] \to A_i \otimes_k k'$ is also finite and injective, so $d = \dim(A_i \otimes_k k')$, as wanted. □

1.2. The Riemann-Roch theorem

Let k be a field and X be a nonsingular projective curve over k.

DEFINITION 1.2.1. The *group of divisors* $\mathrm{Div}(X/k)$ of X is the free abelian group on the set of closed points of X. The *degree* of a divisor D is defined as the sum $\deg D = \sum_x n_x \deg x \in \mathbb{Z}$, where $\deg x = \dim_k k(x)$. The subgroup of divisors of degree zero is denoted by $\mathrm{Div}^0(X/k)$.

Thus a divisor $D \in \mathrm{Div}(X/k)$ is a linear combination of the form $\sum_x n_x x$ where $n_x \in \mathbb{Z}$ and $n_x = 0$ for all but finitely many x. Notice that if k is algebraically closed then $\deg x = 1$ for all x and in particular the degree of D is $\sum_x n_x$.

DEFINITION 1.2.2. We define the partial order on $\mathrm{Div}(X/k)$ given by

$$\sum_x m_x x \geq \sum_x n_x x \quad \text{if and only if} \quad m_x \geq n_x,$$

and we say that $D \in \mathrm{Div}(X/k)$ is an *effective divisor* if $D \geq 0$.

DEFINITION 1.2.3. Let $f \in k(X)$ be nonzero. The *principal divisor* associated to f is the divisor $\mathrm{div}\, f = \sum_x \mathrm{ord}_x(f)x$, which has degree zero by Proposition 1.1.4. We denote by $P(X/k)$ the group of principal divisors.

DEFINITION 1.2.4. For each divisor D we define the k-vector space

$$L(D) = \{f \in k(X) \setminus \{0\} : \operatorname{div} f + D \geq 0\} \cup \{0\},$$

and we denote by $\ell(D)$ its dimension.

Notice that $L(D)$ is indeed a vector space since $\operatorname{ord}_x(f+g) \geq \min(\operatorname{ord}_x(f), \operatorname{ord}_x(g))$.

DEFINITION 1.2.5. Two divisors D and D' are said to be equivalent if $D - D' = \operatorname{div} g$ for some nonzero $g \in k(X)$.

If D and D' are equivalent, then the k-vector spaces $L(D)$ and $L(D')$ are isomorphic via the isomorphism $f \mapsto fg$. Using this remark we can prove that the dimension $\ell(D)$ of $L(D)$ is always finite.

LEMMA 1.2.6. *$L(D)$ is a finite-dimensional k-vector space for all $D \in \operatorname{Div}(X/k)$.*

PROOF. If $L(D)$ is nonzero there is some nonzero f such that $D + \operatorname{div} f \geq 0$, and replacing D with $D + \operatorname{div} f$ we may assume $D \geq 0$. The proof is by induction on the degree of D. If $\deg D = 0$ then $D = 0$, so $L(D) = k$ has dimension one. Now, if $D \neq 0$ write $D = D' + x$ for some point x. If $L(D) = L(D')$ we are done. Otherwise there exists some $g \in L(D) \setminus L(D')$, and then $\operatorname{ord}_x(f) \geq \operatorname{ord}_x(g)$ for all $f \in L(D)$. Hence the map $L(D) \to k(x), f \mapsto f/g + \mathfrak{m}_x$ is well-defined and the kernel is seen to be $L(D')$. Since both $k(x)$ and $L(D')$ are finite-dimension k-vector spaces, so is $L(D)$. □

DEFINITION 1.2.7. The *Picard groups* of X are the quotients

$$\operatorname{Pic}(X/k) = \operatorname{Div}(X/k)/P(X/k) \quad \text{and} \quad \operatorname{Pic}^0(X/k) = \operatorname{Div}^0(X/k)/P(X/k).$$

Our aim now is to sketch the proof of the RIEMANN-ROCH theorem, which will be of utmost importance in the following sections. We follow closely [22].

THEOREM 1.2.8 (RIEMANN-ROCH THEOREM). *There exists an integer $g \geq 0$ and a class in $\operatorname{Pic}(X/k)$ such that for all W in the class and all $D \in \operatorname{Div}(X/k)$ we have*

$$\ell(D) = \deg D - g + 1 + \ell(W - D).$$

DEFINITION 1.2.9. The integer g of Theorem 1.2.8 is called the *genus* of X/k.

COROLLARY 1.2.10. *If W is as in Theorem 1.2.8, then $\ell(W) = g$ and $\deg W = 2g - 2$. Furthermore, if D is a divisor such that $\deg D \geq 2g - 1$, then*

$$\ell(D) = \deg D - g + 1.$$

PROOF. Setting $D = 0$ in Theorem 1.2.8 we get that $\ell(W) = g$, and then putting $D = W$ we deduce that $\deg W = 2g-2$. Finally, if $\deg D \geq 2g-1$ then $\deg(W-D) < 0$, so $\ell(W - D) = 0$ and this proves the corollary. □

To prove Theorem 1.2.8, we first prove the weaker RIEMANN inequality, which says that $\ell(D) \geq \deg D - g + 1$. Once this is proved, it suffices to show that there exists some *canonical* divisor W such that the dimension of $L(W - D)$ coincides with the difference $\ell(D) - (\deg D - g + 1)$, and that is achieved by studying some special linear maps called WEIL differentials, which are defined below. We will associate a unique divisor to each of them, and prove that they form a class in $\mathrm{Pic}(X/k)$ satisfying the required condition.

DEFINITION 1.2.11. Let $K = k(X)$ be the function field of X/k. The *ring of adèles* of X is defined as

$$A_K = \Big\{(a_x)_x \in \prod_{x \text{ closed}} \widehat{K}_x : a_x \in \widehat{\mathcal{O}}_x \text{ for all but finitely many } x\Big\}$$

where $\widehat{K}_x$ is the completion of K with respect to the valuation ord_x and $\widehat{\mathcal{O}}_x$ is the ring of integers of $\widehat{K}_x$. We embed K in A_K via the map $a \mapsto (a)_x$, and we also define for each divisor $D = \sum_x n_x x \in \mathrm{Div}(X/k)$ the subset $A_K(D)$ of points $(a_x)_x \in A_K$ such that $\mathrm{ord}_x(a_x) + n_x \geq 0$ for all x.

Notice that both A_K and $A_K(D)$ are k-vector spaces, and $A_K(D) \cap K = L(D)$.

DEFINITION 1.2.12. A WEIL differential is a linear map $\omega : A_K \to k$ vanishing on K and on $A_K(D)$ for some divisor D. We denote by Ω_K the set of all WEIL differentials and by $\Omega_K(D)$ the set of WEIL differentials vanishing on $A_K(D)$.

Defining $f\omega : g \mapsto \omega(fg)$ for $f \in K$ and $\omega \in \Omega_K$ we see that Ω_K is a K-vector space. Indeed, one checks that if $\omega(A_K(D)) = 0$, then $(f\omega)(A_K(D - \mathrm{div}\, f)) = 0$.

The first step in the proof of the RIEMANN-ROCH theorem is to prove RIEMANN's inequality, which says that there exists a unique integer $g \geq 0$ such that

$$\ell(D) \geq \deg D - g + 1$$

for all divisors D, with equality for $\deg D$ large enough.

LEMMA 1.2.13. *If $D \leq D'$ then $A_K(D) \subseteq A_K(D')$ and*

$$\dim_k(A_K(D')/A_K(D)) = \deg D' - \deg D.$$

PROOF. By induction on $\deg D'$ it suffices to show $\dim_k(A_K(D+x)/A_K(D)) = \deg x$ for all x. Let $n = \operatorname{ord}_x(D)$ and $\widehat{\mathfrak{m}}_x = \mathfrak{m}_x\widehat{\mathcal{O}}_x$. Then the natural map $A_K(D+x) \to \widehat{\mathfrak{m}}_x^{-n}$ induces an isomorphism of k-vector spaces $A_K(D+x)/A_K(D) \simeq \widehat{\mathfrak{m}}_x^{-n-1}/\widehat{\mathfrak{m}}_x^{-n}$, and the claim follows from the isomorphisms $\widehat{\mathfrak{m}}_x^{-n-1}/\widehat{\mathfrak{m}}_x^{-n} \simeq \mathfrak{m}_x^{-n-1}/\mathfrak{m}_x^{-n} \simeq k(x)$. □

Now an application of the isomorphism theorems yields

$$\frac{A_K(D')+K}{A_K(D)+K} \simeq \frac{A_K(D')/A_K(D)}{(A_K(D)+L(D'))/A_K(D)} \simeq \frac{A_K(D')/A_K(D)}{L(D')/L(D)}$$

so the dimension of $(A_K(D')+K)/(A_K(D)+K)$ is $r(D')-r(D)$ by Lemma 1.2.13, where $r(D) = \deg D - \ell(D)$. In particular $r(D') \geq r(D)$.

Fix some transcendental element $f \in K$. We know that $K/k(f)$ is a finite algebraic extension. Let $D_0 = -\sum_{x:\operatorname{ord}_x(f)<0} \operatorname{ord}_x(f)x \in \operatorname{Div}(X/k)$. Then RIEMANN's inequality follows if we prove the following claims:

(1) The increasing sequence $\{r(mD_0)\}_{m\geq 1}$ is bounded.
(2) Every D is equivalent to some divisor D' such that $D' \leq mD_0$ for some m.
(3) RIEMANN's inequality is an equality if $\deg D$ is large enough.

Notice that if g is the smallest integer such that $r(mD_0) \leq g-1$ for all $m \geq 0$, then $-1 = r(0) \leq r(mD_0) \leq g-1$, so $g \geq 0$. To prove (1) we only need to show that the sequence $\ell(mD_0) - m\deg D_0$ is bounded below by some constant, and for this we show that there is a constant m_0 such that for all $m \geq m_0$ we can find $(m-m_0+1)\deg D_0$ linearly independent elements in $L(mD_0)$. We saw in Section 1.1 that $\deg D_0$ is equal to the degree $n = [K : k(f)]$. Let $g_1, \ldots, g_n$ be a $k(f)$-basis of K. We may suppose that these elements belong to the integral closure of $k[f]$ in K. This guarantees that $\operatorname{ord}_x(g_j) \geq 0$ for all x which are not poles of f. Let m_0 be large enough so that $\operatorname{div} g_j + m_0 D_0 \geq 0$ for all j. Then $g_j \in L(mD_0)$ for all $m \geq m_0$, and we claim that the elements

$$\{f^i g_j : 0 \leq i \leq m-m_0, 1 \leq j \leq n\},$$

which are linearly independent over k, belong to $L(mD_0)$. Indeed, if x is a pole of f then

$$\operatorname{ord}_x(f^i g_j) = i\operatorname{ord}_x(f) + \operatorname{ord}_x(g_j) \geq (m-m_0)\operatorname{ord}_x(f) + m_0\operatorname{ord}_x(f) = m\operatorname{ord}_x(f).$$

Hence we get the desired inequality $r(mD_0) \leq (m-m_0+1)\deg D_0$.

To prove (2) we need to find some $h \in K^\times$ such that $mD_0 + \operatorname{div} h - D \geq 0$. For each pole x of $-D$ which is not a pole of f let h_x be the generator of the maximal ideal $k[f] \cap \mathfrak{m}_x$. Then letting h be the product of these h_x's with sufficiently large exponents

we have that the poles of $\operatorname{div} h - D$ occur among the poles of f. Letting m be large enough (2) follows.

Finally we prove (3). Let m_1 large so that $r(m_1 D_0) = g-1$. Then $\ell(D - m_1 D_0) \geq 1$ if $\deg D \geq m_1 \deg D_0 + g$, so there exists some nonzero $h \in L(D - m_1 D_0)$. Hence $m_1 D_0 \leq D + \operatorname{div} h$, so $g - 1 = r(m_1 D_0) \leq r(D)$.

The next step is to analyse the difference $\ell(D) - (\deg D - g + 1)$. In the following lemma we prove that it is exactly the dimension of the k-vector space $\Omega_K(D)$. Then we will show that $\Omega_K(D)$ coincides with $\ell(W - D)$ for some appropriate W, and the Riemann-Roch theorem will follow easily.

Lemma 1.2.14. *For any divisor D the vector space $\Omega_K(D)$ is finite-dimensional and*

$$\ell(D) = \deg D - g + 1 + \dim_k \Omega_K(D).$$

In particular $g = \dim_k \Omega_K(0)$.

Proof. Fix a divisor $D_0 \geq D$ with degree large enough so that Riemann's inequality holds. Then for all $D' \geq D_0$ the computation above shows that

$$\dim_k \frac{A_K(D') + K}{A_K(D) + K} = g - 1 + \ell(D) - \deg D.$$

In particular $A_K(D') + K = A_K(D_0) + K$ for all $D' \geq D_0$. Since for all $\xi \in A_K$ we can find some D' such that $\xi \in A_K(D')$ we must have $A_K(D_0) + K = A_K$. But $\Omega_K(D)$ is the k-dual space of $A_K/(A_K(D) + K)$, so the lemma follows. □

We can associate to each nonzero differential $\omega \in \Omega_K$ a unique divisor $\operatorname{div} \omega$ such that $\omega \in \Omega_K(\operatorname{div} \omega)$ and with the following property: if $\omega \in \Omega_K(D)$ for some divisor D then $D \leq \operatorname{div} \omega$. Notice that if $\deg D$ is large enough then we showed in the proof of Lemma 1.2.14 that $A_K = A_K(D) + K$, so for any such D we have $\omega(A_K) = 0$, this is, $\omega = 0$. Hence the divisors D with $\omega \in \Omega_K(D)$ have bounded degrees, so there exists some $\operatorname{div} \omega = \sum_x n_x x$ with maximal degree, and the claim follows if we prove that if $D = \sum_x n'_x x$ is such that $\omega \in \Omega_K(D)$, then $D' = \sum_x \max(n_x, n'_x) x$ also satisfies $\omega \in \Omega_K(D')$. But one checks that $A_K(\operatorname{div} \omega) + A_K(D) = A_K(D')$.

Lemma 1.2.15. *If $f \in K^\times$ and $\omega \in \Omega_K$ then* $\operatorname{div}(f\omega) = \operatorname{div} f + \operatorname{div} \omega$.

Proof. Since $f\omega$ vanishes on $A_K(\operatorname{div} f + \operatorname{div} \omega)$ we have $\operatorname{div} f + \operatorname{div} \omega \leq \operatorname{div}(f\omega)$. Similarly $\operatorname{div} f^{-1} + \operatorname{div}(f\omega) \leq \operatorname{div} \omega$, and the lemma follows from $\operatorname{div} f + \operatorname{div} f^{-1} = 0$. □

By Theorem 1.2.16 below any two nonzero differentials differ by some $f \in K^\times$, and therefore by Lemma 1.2.15 the set $\operatorname{div}(\Omega_K \setminus \{0\})$ defines a unique canonical class in $\operatorname{Pic}(X/k)$. We also see that $L(\operatorname{div}\omega - D)\omega \subseteq \Omega_K(D)$ since given $f \in L(\operatorname{div}\omega - D)$ we have $\operatorname{div}(f\omega) = \operatorname{div} f + \operatorname{div}\omega \geq D$, so $A_K(D) \subseteq A_K(\operatorname{div}(f\omega))$. The reverse inclusion also holds: if $\omega' \in \Omega_K(D)$ then $\omega' = f\omega$ for some nonzero f and $D \leq \operatorname{div}\omega' = \operatorname{div} f + \operatorname{div}\omega$. This implies that $\operatorname{div} f \in L(\operatorname{div}\omega - D)$, so $\operatorname{div}\omega' \in L(\operatorname{div}\omega - D)\omega$. Hence we have an isomorphism $L(\operatorname{div}\omega - D) \simeq \Omega_K(D)$ of k-vector spaces. This readily implies the Riemann-Roch Theorem 1.2.8.

Theorem 1.2.16. $\dim_K \Omega_K = 1$.

Proof. We saw above that $L_\omega = L(\operatorname{div}\omega - D)\omega$ is a subspace of $\Omega_K(D)$. Thus the theorem follows if we prove that L_ω and $L_{\omega'}$ have nonzero intersection for some suitable D. Fix some point x and put $D = -nx$ for n large. Then by Lemma 1.2.14 we have $\dim_k \Omega_K(-nx) = n\deg x + g - 1$. On the other hand, using Riemann's inequality,

$$\dim_k L_\omega + \dim_k L_{\omega'} \geq 2n\deg x + \deg\omega + \deg\omega' - 2g + 2.$$

Thus, if n is large enough, the sum of the dimensions of L_ω and $L_{\omega'}$ is greater than the dimension of $\Omega_K(D)$, so they must have nonzero intersection. □

1.3. The zeta function of a curve

Let $k = \mathbb{F}_q$ be a finite field of q elements, let $k_r = \mathbb{F}_{q^r}$ for each $r \geq 1$, and let X/k be a geometrically irreducible projective nonsingular curve of genus g. Let $X_r = X \times_k \operatorname{Spec} k_r$ and denote by ν_r the number of k_r-rational points of X. We introduce the *zeta function* of X/k as the formal series

$$Z(X/k, T) = \exp\left(\sum_{r=1}^{\infty} \nu_r T^r/r\right). \tag{1.3.1}$$

We also use the notation $\zeta(X/k, s) = Z(X/k, q^{-s})$. The following theorem was conjectured by E. Artin in 1924 and proved by A. Weil in the 1940s, who then conjectured generalisations to arbitrary algebraic varieties over finite fields.

Theorem 1.3.2 (Weil conjectures for algebraic curves). *Let X/k be a geometrically irreducible projective nonsingular curve of genus g over a finite field k of q elements. Then:*

(Z1) RATIONALITY: *$Z(X/k,T)$ is a rational function. More precisely, there exist algebraic integers $\omega_1,\dots,\omega_{2g}$ such that*

$$Z(X/k,T)=\frac{(1-\omega_1 T)\cdots(1-\omega_{2g}T)}{(1-T)(1-qT)}.$$

(Z2) FUNCTIONAL EQUATION: *$Z(X/k,T)$ satisfies the functional equation*

$$\xi(X/k,s)=\xi(X/k,1-s)\quad \textit{where}\quad \xi(X/k,s)=q^{(g-1)s}\zeta(X/k,s).$$

Equivalently, $Z(X/k,1/qT)=(qT^2)^{1-g}Z(X/k,T)$.

(Z3) RIEMANN HYPOTHESIS: *$|\omega_i|=q^{\frac{1}{2}}$ for all $1\le i\le 2g$. In other words, the zeros of $Z(X/k,T)$ all lie on the critical line $\operatorname{Re} s=\frac{1}{2}$.*

The first two statements are direct applications of the RIEMANN-ROCH theorem and are proved in this section. In Section 1.4 we prove the RIEMANN hypothesis following BOMBIERI [3], who gave a simple proof based on an elementary but powerful method introduced by S. A. STEPANOV. The original form of this method is examined more closely in Chapter 2.

To study the zeta function $Z(X/k,T)$ it is convenient to rewrite it in a different way. The following general construction is motivated by algebraic number theory. Let Y be a scheme of finite type over $\mathbb{Z}$ and define the zeta function $\zeta_Y(s)=\prod_y(1-|k(y)|^{-s})^{-1}$, where the product runs over all closed points of Y. When $Y=\operatorname{Spec}\mathcal{O}_L$ is the spectrum of the ring of integers of a number field L it coincides with the usual DEDEKIND zeta function. Now assume further that Y is also a k-scheme. Then it is easy to see that $\zeta_Y(s)=\sum_{d\ge1}b_dq^{-ns}$, where b_d is the number of effective divisors of degree d, and furthermore if a_d is the number of points of degree d then we can write the above EULER product in the form $\zeta_Y(s)=\prod_{d\ge1}(1-q^{-sd})^{-a_d}$.

LEMMA 1.3.3. *The zeta function $\zeta_X(s)$ coincides with the zeta function $\zeta(X/k,s)$ associated to the curve X/k.*

PROOF. Taking logarithms in the Euler product of $\zeta_X(s)$ we have

$$\log\zeta_X(s)=-\sum_{d\ge1}a_d\log(1-q^{-sd})=\sum_{n\ge1}\sum_{d\ge1}a_dq^{-snd}/n=\sum_{r\ge1}q^{-sr}\sum_{dn=r}a_d/n$$

and comparing this series with the definition (1.3.1) of $\zeta(X/k,s)$ the claim will follow if we prove that $\nu_r=\sum_{d|r}da_d$. Indeed, notice that $\operatorname{Hom}_k(k_d,k_r)$ has cardinality d if d divides r and it is empty otherwise. Thus, since there is a one-to-one correspondence

between k_r-rational points and pairs (x, ι) with x closed in X and $\iota \in \mathrm{Hom}_k(k(x), k_r)$, where $k(x) = k_{\deg x}$, we have

$$|X(k_r)| = \sum_{d \geq 1} \#\{(x, \iota) : x \text{ closed in } X, \iota \in \mathrm{Hom}_k(k_d, k_r), \deg x = d\} = \sum_{d|r} d a_d$$

as a_d is the number of closed points of degree d. □

We are ready to prove the rationality of $Z(X/k, T)$. First we compute explicitly the number of effective divisors in a given class of $\mathrm{Pic}(X/k)$.

LEMMA 1.3.4. *Let $D \in \mathrm{Div}(X/k)$. Then the number of effective divisors in the class of D is exactly $(q^{\ell(D)} - 1)/(q - 1)$.*

PROOF. Clearly the class of D contains effective divisors if and only if $\ell(D) > 0$, so it suffices to show the formula in this case. The map from $L(D) \setminus \{0\}$ to the set of effective divisors in the class of D sending $\mathrm{div}\, f$ to $\mathrm{div}\, f + D$ is clearly surjective. On the other hand, $\mathrm{div}\, f = \mathrm{div}\, f'$ if and only if $f' f^{-1} \in k^\times$. Since $L(D) \setminus \{0\}$ has $q^{\ell(D)} - 1$ elements and $k^\times$ has $q - 1$ elements the result follows. □

PROOF OF THEOREM 1.3.2(Z1). Let $h_X^0 = |\mathrm{Pic}^0(X/k)|$ be the number of classes of divisors of degree zero. The degree map $\deg : \mathrm{Pic}(X/k) \to \mathbb{Z}$ has image $m\mathbb{Z}$ for some positive integer m. Later we will prove that in fact $m = 1$. Since the kernel of deg is precisely $\mathrm{Pic}^0(X/k)$, we have that the number of classes of divisors of degree d is h_X^0 if d is multiple of m and zero otherwise. Thus by the RIEMANN-ROCH formula and Lemma 1.3.4 we have for $d \geq 2g - 1$ that $b_d = h_X^0 (q^{d-g+1} - 1)(q - 1)^{-1}$ if m divides d and $b_d = 0$ otherwise. Hence, letting n_0 be the smallest integer such that $n_0 m \geq 2g - 1$, after splitting the series in two parts we get

$$Z(X/k, T) = \sum_{n=0}^{n_0 - 1} b_{nm} T^{nm} + h_X^0 T^{n_0 m} f(T^m)(1 - T^m)^{-1}(1 - (qT)^m)^{-1}$$

where $f(T)$ is the polynomial with integer coefficients given by

$$f(T) = (q - 1)^{-1}(q^a - 1 - T(q^a - q^m)), \quad a = n_0 m - g + 1.$$

In particular, we see that the proof of (Z1) is reduced to showing that $m = 1$. This will follow from the following lemma, which relates the zeta function of the curve X_n and the zeta function of X.

LEMMA 1.3.5. *If $Z(X_n/k_n, T)$ is the zeta function of the curve X_n, then it satisfies*

$$Z(X_n/k_n, T^n) = \prod_{\zeta^n = 1} Z(X/k, \zeta T) \tag{1.3.6}$$

where the product runs over all nth roots of unity.

PROOF. If ν'_r is the number of k_{rn}-rational points of X_n then clearly $\nu'_r = \nu_{rn}$, so

$$\sum_{r=1}^{\infty} \nu'_r T^{rn}/r = \sum_{r=1}^{\infty} \nu_{rn} T^{rn}/r = \sum_{\zeta^n=1} \sum_{r=1}^{\infty} \nu_r (\zeta T)^r/r$$

and taking the exponential yields (1.3.6). □

Notice that, since $f(1) = 1 + q + \cdots + q^{m-1}$ is nonzero, the function $Z(X/k, T)$ has a simple pole at $T = 1$. Thus, putting $n = m$ in Equation (1.3.6) we have that $Z(X_m/k_m, T^m) = (Z(X/k, T))^m$, and comparing the order of the poles at $T = 1$ we see that $m = 1$, as wanted. □

PROOF OF THEOREM 1.3.2(Z2). Using Lemma 1.3.4 and arguing as above we can write $Z(X/k, T) = (q-1)^{-1}(A(T) + h^0_X B(T))$ where

$$A(T) = \sum_{0 \le \deg D \le 2g-2} q^{\ell(D)} T^{\deg D}$$

$$B(T) = \sum_{n=2g-1}^{\infty} q^{n-g+1} T^n - \sum_{n=0}^{\infty} T^n = \frac{q^g T^{2g-1}}{1-qT} - \frac{1}{1-T}.$$

It is easy to see now that $B(1/qT) = q^{1-g} T^{2-2g} B(T)$, so to prove (Z2) it suffices to show that the same relation holds for $A(T)$. Let W be a canonical divisor, which has degree $2g-2$. Since the map $D \mapsto W - D$ induces a bijection between the classes of divisors of degree n and the classes of divisors of degree $2g-2-n$ for each $0 \le n \le 2g-2$, we have

$$A(T) = \sum_{0 \le \deg D \le 2g-2} q^{\ell(W-D)} T^{\deg(W-D)} = T^{\deg W} \sum_{0 \le \deg D \le 2g-2} q^{\ell(D)-1+g-\deg D} T^{-\deg D}$$

on using the RIEMANN-ROCH theorem. Hence $A(T) = q^{g-1} T^{2g-2} A(1/qT)$ and the proof of (Z2) is finished. □

To end the section, we remark that the functional equation can be rewritten in the form

$$\prod_{i=1}^{2g} (qT - \omega_i) = q^{2g} \prod_{i=1}^{2g} (1 - \omega_i T).$$

In particular $\omega_1 \cdots \omega_{2g} = q^g$, so they are nonzero algebraic numbers, and the map $\omega_i \mapsto q/\omega_i$ is a bijection of the set $\{\omega_1, \ldots, \omega_{2g}\}$. Furthermore, from the rationality it follows at once the identity

$$\nu_r = q^r - \sum_{i=1}^{2g} \omega_i^r + 1, \quad \text{for all } r \ge 1.$$

1.4. Bombieri's proof of the Riemann hypothesis

Keep the notation of the previous section. Our goal now is to prove the RIEMANN hypothesis (Z3), this is, that $|\omega_i| = q^{\frac{1}{2}}$ for all $1 \leq i \leq 2g$. We start with a couple of remarks. Let $\omega'_1, \dots, \omega'_{2g}$ be the numbers associated with the curve X_r. Then from Equation (1.3.6) we have

$$Z(X_r/k_r, T^r) = \prod_{\zeta^r=1} \frac{(1-\omega_1\zeta T)\cdots(1-\omega_{2g}\zeta T)}{(1-\zeta T)(1-q\zeta T)} = \frac{(1-\omega_1^r T^r)\cdots(1-\omega_{2g}^r T^r)}{(1-T^r)(1-q^r T^r)},$$

so, up to a rearrangement, $\omega'_i = \omega_i^r$. In particular, the RIEMANN hypothesis holds for X_r if and only if it holds for X. Thus we may assume that q is a square and large enough, which will be convenient in the proof. On the other hand, it is enough to prove that $\nu_r = q^r + O(q^{\frac{r}{2}})$, the implied constant independent of r. Indeed, this is consequence of the following general lemma applied to the numbers $\omega_1, \dots, \omega_{2g}$, which implies that $|\omega_i| \leq \sqrt{q}$ for all i. But $\omega_i\omega_j = q$ for some j, so this is enough to deduce that $|\omega_i| = \sqrt{q}$.

LEMMA 1.4.1. *Let $z_1, \dots, z_m$ be complex numbers and suppose that there exists some constants A and B such that the inequality*

$$|z_1^n + \cdots + z_m^n| \leq AB^n$$

holds for all n large enough. Then $|z_i| \leq B$ for all $1 \leq i \leq m$.

PROOF. The radius of convergence $\min_i |z_i|^{-1}$ of

$$\sum_{r\geq 1}\sum_i z_i^r X^r = \sum_i (1-z_iX)^{-1}$$

is at least B^{-1} by the hypothesis, so we must have $|z_i| \leq B$ for all i. □

Fix an algebraic closure $\overline{k}$ of k and let $\overline{X} = X \times_k \operatorname{Spec}\overline{k}$. We know that the number of k-rational points is exactly the number of closed points of $\overline{X}$ fixed by the arithmetic FROBENIUS ϕ_a of $\overline{X}$. The idea in STEPANOV's method is to construct a rational function $f \in k(\overline{X})$ that vanishes at all but some fixed set of $\overline{k}$-rational points with order of vanishing at least some big integer m. Thus, if there are m_0 exceptions, f has at least $m(\nu_1 - m_0)$ zeros, and since f has as many zeros as poles we have the estimation

$$\nu_1 \leq m_0 + \frac{1}{m}(\#\text{ poles of } f). \tag{1.4.2}$$

Therefore, if the number of poles of f is not very large we obtain a good upper bound for ν_1. Indeed, we will prove under suitable hypotheses that $\nu_1 \leq q + O(q^{\frac{1}{2}})$. Then with

a Galois-theoretic argument we will show that the reverse inequality also holds, and consequently $\nu_1 = q+O(q^{\frac{1}{2}})$. Finally by base change to k_r we get the desired estimation for ν_r.

1.4.1. The upper bound. Let p be the characteristic of k and write $q = p^\alpha$. The main theorem we prove is the following.

Theorem 1.4.3. *If α is even and $q > (g+1)^4$ then $\nu_1 < q + (2g+1)q^{\frac{1}{2}} + 1$.*

Before proving the theorem we introduce the following notation. Since the inequality is trivial if ϕ_a has no fixed points, we assume that ν_1 is nonzero and we fix a point x_0 fixed by ϕ_a. Consider the $\overline{k}$-vector space $R_m = L(mx_0)$ whose nonzero elements are nonzero rational functions on $\overline{X}$ with at most a pole of order m at x_0. By the Riemann-Roch theorem we have $\dim R_m \geq m+1-g$ with equality if $m \geq 2g-1$. Since $W-(m+1)x_0 \leq W-mx_0$ for a canonical divisor W, we also see that $\dim R_m \leq \dim R_{m+1} \leq \dim R_m + 1$. In particular induction shows that $\dim R_m \leq m+1$ since $\dim R_0 = 1$. Also notice that if $f \in R_m$ then $f^p \in R_{mp}$ since $\operatorname{div} f^p + mpx_0 = p(\operatorname{div} f + mx_0)$. We denote by $R_\ell^{(p^\mu)}$ the subspace of $R_{\ell p^\mu}$ consisting of p^μ-powers of elements in R_ℓ, which is isomorphic to R_ℓ via the isomorphism $R_\ell \to R_\ell^{(p^\mu)}, f \mapsto f^{p^\mu}$. Finally we let $(\phi_r^F)^* R_m$ be the subspace consisting of elements of the form $(\phi_r^F)^*(f)$ with $f \in R_m$. Since $\phi_r^F = \phi^F \circ \phi_g^F$, we have

$$(\phi_r^F)^*(f) = ((\phi^F)^* \circ (\phi_g^F)^*)(f) = ((\phi_g^F)^*(f))^q$$

so all the elements in $(\phi_r^F)^* R_m$ are qth powers, and furthermore $(\phi_r^F)^* R_m \subseteq R_{qm}$ since

$$\operatorname{ord}_{x_0}((\phi_r^F)^*(f)) = q \operatorname{ord}_{x_0}((\phi_g^F)^*(f)) = q \operatorname{ord}_{\phi_g^F(x_0)}(f) = q \operatorname{ord}_{x_0}(f).$$

If A is a subspace of R_m and B is a subspace of R_n, we let AB be the subspace of R_{mn} spanned by products of the form fg with $f \in R_m$ and $g \in R_n$. A crucial preliminary result in the proof of Theorem 1.4.3 is the following lemma.

Lemma 1.4.4. *For $\ell p^\mu < q$ the natural map $R_\ell^{(p^\mu)} \otimes_{\overline{k}} ((\phi_r^F)^* R_m) \to R_\ell^{(p^\mu)}((\phi_r^F)^* R_m)$ is an isomorphism.*

Proof. The map is clearly onto so we only need to prove that it is injective. Since $\dim R_{m+1} \leq \dim R_m + 1$ we can find a basis $s_1, \ldots, s_r$ of R_m such that

$$\operatorname{ord}_{x_0}(s_1) < \operatorname{ord}_{x_0}(s_2) < \cdots < \operatorname{ord}_{x_0}(s_r).$$

Let $g_i \in R_\ell$ be such that $\sum_{i=1}^r g_i^{p^\mu}(\phi_r^F)^*(s_i) = 0$, and suppose that $g_1 = \cdots = g_{\rho-1} = 0$ but $g_\rho \neq 0$. Then

$$\operatorname{ord}_{x_0}(g_\rho^{p^\mu}(\phi_r^F)^*(s_\rho)) = \operatorname{ord}_{x_0}\Big(\sum_{i=\rho+1}^r g_i^{p^\mu}(\phi_r^F)^*(s_i)\Big) \geq \min_{i\geq\rho+1} \operatorname{ord}_{x_0}(g_i^{p^\mu}(\phi_r^F)^*(s_i))$$

and since $\operatorname{ord}_{x_0}(g_i) \geq -\ell$ and $\operatorname{ord}_{x_0}(s_i) \geq \operatorname{ord}_{x_0}(s_{\rho+1})$ for $i \geq \rho+1$ we have

$$p^\mu \operatorname{ord}_{x_0}(g_\rho) \geq -\ell p^\mu + q(\operatorname{ord}_{x_0}(s_{\rho+1}) - \operatorname{ord}_{x_0}(s_\rho)) \geq -\ell p^\mu + q,$$

which is positive by hypothesis. This means that g_ρ vanishes at x_0, so it must be identically zero since it has no poles outside x_0. □

Proof of Theorem 1.4.3. Assume that $\ell p^\mu < q$, so that by Lemma 1.4.4 and the remarks preceding it follows that $\dim R_\ell^{p^\mu}((\phi_r^F)^* R_m) = \dim R_\ell \dim R_m$. Keeping the notation used in the proof of the lemma, we also have that the map

$$\delta : R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) \to R_\ell^{(p^\mu)} R_m,\ \sum_{i=1}^r g_i^{p^\mu}(\phi_r^F)^*(s_i) \mapsto \sum_{i=1}^r g_i^{p^\mu} s_i$$

is well-defined as it is the composition of $R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) \to R_\ell^{(p^\mu)} \otimes_{\overline{k}} ((\phi_r^F)^* R_m)$ and the natural map $R_\ell^{(p^\mu)} \otimes_{\overline{k}} ((\phi_r^F)^* R_m) \to R_\ell^{(p^\mu)} R_m$.

Suppose that the kernel of δ contains a nonzero $f = \sum_i g_i^{p^\mu}(\phi_r^F)^*(s_i)$. Notice that, if x is a fixed point of ϕ_a, or equivalently a fixed point of ϕ_r, then $(\phi_r^F)^*(g) - g \in \mathfrak{m}_x$ for all $g \in k(X)$ regular at x. Indeed, we know that the natural inclusion $\overline{k} \to k(x)$ is an isomorphism, so if $\operatorname{Spec} A \subseteq X$ is an affine neighbourhood of x and $\mathfrak{p}$ is the prime of A corresponding to x, then the diagram

$$\begin{array}{ccccc} & A & \longrightarrow & A_{\mathfrak{p}} & \twoheadrightarrow k(x) \\ \overline{k} \nearrow & \downarrow{\scriptstyle \phi_r^\#} & & \downarrow{\scriptstyle (\phi_r^\#)_{\mathfrak{p}}} & \| \\ \overline{k} \searrow & A & \longrightarrow & A_{\mathfrak{p}} & \twoheadrightarrow k(x) \end{array}$$

commutes, so g and $\phi_r^\#(g)$ have the same image in $k(x)$. Therefore, for all fixed points $x \neq x_0$ of ϕ_a we have

$$\operatorname{ord}_x(f) = \operatorname{ord}_x\Big(\sum_{i=1}^r g_i^{p^\mu}((\phi_r^F)^*(s_i) - s_i)\Big) > 0,$$

this is, x is a zero of f. Since f is a p^μ-power, its zeros have order of vanishing at least p^μ since, so f has at least $p^\mu(\nu_1 - 1)$ zeros. On the other hand, f has at most $\ell p^\mu + mq$

poles as $f \in R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) \subseteq R_{\ell p^\mu + mq}$. Equation (1.4.2) now says that

$$\nu_1 \leq 1 + \frac{1}{p^\mu}(\ell p^\mu + mq) = 1 + \ell + \frac{mq}{p^\mu}$$

and we have to find suitable m, ℓ and μ such that $\ell p^\mu < q$ and $\dim \ker \delta > 0$. Assume $\ell, m \geq g$, so that in particular $\ell p^\mu + m \geq 2g - 1$. Then

$$\dim R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) = \dim R_\ell \dim R_m \geq (\ell + 1 - g)(m + 1 - g)$$
$$\dim R_\ell^{(p^\mu)} R_m \leq \dim R_{\ell p^\mu + m} = \ell p^\mu + m + 1 - g$$

so

$$\begin{aligned}\dim \ker \delta &\geq \dim R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) - \dim R_\ell^{(p^\mu)} R_m \\ &\geq (\ell + 1 - g)(m + 1 - g) - (\ell p^\mu + m + 1 - g) \\ &= (\ell - g)(m + 1 - g) - \ell p^\mu.\end{aligned}$$

Put $\mu = \alpha/2$ and $m = p^\mu + 2g \geq g$. Then $\dim \ker \delta > 0$ if

$$\ell > \frac{g}{g+1} p^\mu + g.$$

Let ℓ be the smallest integer satisfying this inequality. We need to verify that $\ell p^\mu < q$, this is, $\ell < p^\mu$. But if $(g+1)^4 < q = p^\alpha$ then

$$\ell \leq \frac{g}{g+1} p^\mu + g + 1 = p^\mu - \frac{p^\mu - (g+1)^2}{g+1} < p^\mu,$$

and this concludes the proof of Theorem 1.4.3. □

Fix a transcendental element $t \in k(X)$ and consider the finite algebraic extension $k(X)/k(t)$, which corresponds to a surjective morphism $X \to \mathbb{P}^1_k$ of curves. Let σ^{-1} be an automorphism in $\mathrm{Aut}_{k(t)}(k(X))$ and let ϕ_σ be the corresponding morphism $X \to X$ of k-schemes (in fact of $\mathbb{P}^1_k$-schemes) induced by σ^{-1}. Also define the set

$$X(k, \sigma) = \{x \in \overline{X}(\overline{k}) : \phi_a^F(x) = \phi_{\sigma,r}(x)\}$$

where $\phi_{\sigma,r} = \phi_\sigma \times_k \mathrm{id}_{\mathrm{Spec}\,\overline{k}}$. When σ is the identity this is exactly $X(k)$, the set of k-rational points on X. Let $\nu_{1,\sigma} = |X(k,\sigma)|$ and fix some $x_0 \in X(k,\sigma)$, if it exists. Then we have the following generalisation of Theorem 1.4.3.

THEOREM 1.4.5. *If α is even and $q > (g+1)^4$ then $\nu_{1,\sigma} < q + (2g+1)q^{\frac{1}{2}} + 1$.*

PROOF. We only need to replace δ with the map

$$\delta_\sigma : R_\ell^{(p^\mu)}((\phi_r^F)^* R_m) \to R_\ell^{(p^\mu)}(\phi_{\sigma,r}^* R_m),\ \sum_{i=1}^r g_i^{p^\mu}(\phi_r^F)^*(s_i) \mapsto \sum_{i=1}^r g_i^{p^\mu}\phi_{\sigma,r}^*(s_i).$$

and argue as in the proof of Theorem 1.4.3. □

1.4.2. The lower bound. In Section 1.4.1 we proved that $\nu_1 < q + (2g+1)q^{\frac{1}{2}} + 1$, so by the discussion at the beginning of Section 1.4 we are left with proving the reverse inequality $\nu_1 > q + O(q^{\frac{1}{2}})$. This will follow from Theorem 1.4.5 and a GALOIS-theoretic trick. As at the end of the previous section let $\phi : X \to \mathbb{P}^1_k$ be a morphism of curves. While in general it is not a GALOIS cover, by the following lemma we may assume it is separable, this is, the extension $k(X)/\phi^* k(\mathbb{P}^1_k)$ of function fields is separable.

LEMMA 1.4.6. *Let k be a field of characteristic p and K be a finitely generated field of transcendence degree one over k. Then there exists $t \in K$ such that $K/k(t)$ is finite and separable.*

PROOF. Let $E \subseteq K$ be a subfield with $[E : K]$ minimal among the fields for which there is some $t \in K$ such that $E/k(t)$ is separable, and suppose that $E \neq K$. Since K/E is purely inseparable there is some $f \in K \setminus E$ such that $f^p \in E$. We claim that $E(f)$ is separable over $k(f)$, contradicting the minimality of E. Let $P(Z,T) \in k[Z,T]$ be a polynomial such that $F(Z,t) \in k(t)[Z]$ is irreducible and $F(f^p,t) = 0$. Since $f^p \in E$ we have that $F(Z,t)$ is separable. In particular not all the coefficients of $F(Z,t)$ lie in $k(t^p)$ since otherwise $P(Z,t) = Q(Z,t^p)$ for some polynomial $Q(Z,T)$, and so f is a root of the separable polynomial $Q(Z,t)$. Hence the polynomial $P(f,T) \in k(f)[T]$ is separable over $k(f)$, so $k(t)$ is separable over $k(f)$. This implies that $E(f)$ is separable over $k(f)$, as wanted. □

Hence there exists a finite GALOIS closure $k(Y)$ of $k(X)/k(t)$, which corresponds to some curve Y/k of genus g'. We also may assume Y is geometrically irreducible and that $q > (g'+1)^4$ after base changing to a finite extension of $k(Y) \cap \overline{k}$ if necessary. Let $\overline{Y} = Y_{\overline{k}}$ and $G = \mathrm{Gal}(k(Y)/k(t))$. Since k is algebraically closed in $k(Y)$ we see that G is naturally isomorphic to the GALOIS group of the extension $k(\overline{Y})/\overline{k}(t)$. Consider the

following commutative diagram:

$$\begin{array}{ccc} \overline{Y} & \xrightarrow{\phi_a^F} & \overline{Y} \\ {\scriptstyle\overline{\psi}}\downarrow & & \downarrow{\scriptstyle\overline{\psi}} \\ \mathbb{P}^1_{\overline{k}} & \xrightarrow[\phi_a^F]{} & \mathbb{P}^1_{\overline{k}} \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\pi} \\ \mathbb{P}^1_k & = & \mathbb{P}^1_k \end{array}$$

If $y \in \overline{Y}$ is a closed point whose image in $\mathbb{P}^1_k$ is k-rational then $\phi_a^F\overline{\psi}(y) = \overline{\psi}(y)$, so there exists some automorphism σ^{-1} such that $\phi_a^F(y) = \phi_{\sigma,r}(y)$ since the Galois-action is transitive. Furthermore, if y is unramified over $\mathbb{P}^1_{\overline{k}}$ then such a σ^{-1} is unique since there are $|G|$ points lying above $\overline{\psi}(y)$. This σ^{-1} is called the Frobenius substitution of y. Define the set

$$A = \{y \in \overline{Y}(\overline{k}) : \pi\overline{\psi}(y) \in \mathbb{P}^1_k(k),\ y \text{ unramified over } \mathbb{P}^1_{\overline{k}}\},$$

which can be decomposed as the disjoint union of the sets

$$A_\sigma = \{y \in \overline{Y}(\overline{k}) : \pi\overline{\psi}(y) \in \mathbb{P}^1_k(k),\ y \text{ unramified over } \mathbb{P}^1_{\overline{k}},\ \phi_a^F(y) = \phi_{\sigma,r}(y)\}$$

with $\sigma \in G$. Since $A_\sigma \subseteq Y(k,\sigma)$, by Theorem 1.4.5 we have $|A_\sigma| \leq q + O(q^{\frac{1}{2}})$ for all $\sigma \in G$. Thus,

$$|A_\sigma| = |A| - \sum_{\tau \neq \sigma} |A_\tau| \geq |A| - q(|G| - 1) + O(q^{\frac{1}{2}}).$$

On the other hand, for each $w \in \mathbb{P}^1_k(k)$ there is exactly one point w' on $\mathbb{P}^1_{\overline{k}}$ lying above w, and if w' is unramified in $\overline{Y}$ then there are $|G|$ points on $\overline{Y}$ lying above w'. Hence

$$|A| = |\mathbb{P}^1_k(k)||G| + O(1) = (q+1)|G| + O(1)$$

where $O(1)$ is bounded by $|G|$ times the number of points on $\overline{Y}$ ramified over $\mathbb{P}^1_{\overline{k}}$, which is finite by a standard result on Dedekind domains (see for example Theorem 7.3, Chapter I in [14]) and does not depend on q. Combining the last two equations we deduce that $|A_\sigma| \geq q + O(q^{\frac{1}{2}})$ for all $\sigma \in G$.

Finally, the map $\psi' : Y \to X$ is a Galois cover with Galois group some subgroup H of G. Notice that if $\sigma \in H$ and $y \in A_\sigma$, then y is unramified over $\overline{X}$, and since $\overline{X}(\overline{k}) = \overline{Y}(\overline{k})^H$ we have

$$\overline{\psi}'(y) = \phi_{\sigma,r}\overline{\psi}'(y) = \overline{\psi}'\phi_{\sigma,r}(y) = \overline{\psi}'\phi_a^F(y) = \phi_a^F\overline{\psi}'(y),$$

this is, $\overline{\psi}'(y) \in \overline{X}$ is fixed by the FROBENIUS, so it is a k-rational point on X. Thus we have proved that $A_\sigma \subseteq A_{\sigma,X}$, where

$$A_{\sigma,X} = \{y \in \overline{Y}(\overline{k}) : \pi_X \overline{\psi}'(y) \in X(k),\ y \text{ unramified over } \overline{X},\ \phi_a^F(y) = \phi_{\sigma,r}(y)\}.$$

Arguing as above with $\mathbb{P}^1_{\overline{k}}$ replaced by $\overline{X}$ we find that

$$\sum_{\sigma \in H} |A_{\sigma,X}| = |X(k)||H| + O(1) = \nu_1 |H| + O(1)$$

where the constant only depends on the number of points on $\overline{X}$ ramified in $\overline{Y}$, whence

$$\nu_1 = \frac{1}{|H|} \sum_{\sigma \in H} |A_{\sigma,X}| + O(1) \geq \frac{1}{|H|} \sum_{\sigma \in H} |A_\sigma| + O(1) \geq q + O(q^{\frac{1}{2}}),$$

where the implied constant only depends on the genus of Y and the number of points on $\mathbb{P}^1_{\overline{k}}$ ramified in $\overline{Y}$ and is independent of q. This inequality combined with Theorem 1.4.3 yields $\nu_1 = q + O(q^{\frac{1}{2}})$. By base change we also have $\nu_r = q^r + O(q^{\frac{r}{2}})$ for all $r \geq 1$, and this concludes the proof of the RIEMANN hypothesis since the constant does not depend on r.

CHAPTER 2

Kloosterman and Weil sums

2.1. Exponential sums over finite fields

2.1.1. Characters. Let $\mathbb{F}_q$ be a finite field of q elements.

Definition 2.1.1. An *additive* (resp. *multiplicative*) character is group homomorphism from the additive group $\mathbb{F}_q^+$ (resp. multiplicative group $\mathbb{F}_q^\times$) to the multiplicative group of nonzero complex numbers $\mathbb{C}^\times$.

Since $\mathbb{F}_q$ is finite, the image of any character is contained in the group μ_m of mth roots of unity, where $m = q$ for additive characters and $m = q - 1$ for multiplicative characters. More generally, a character on a group G is a group homomorphism $\psi \in \mathrm{Hom}(G, \mathbb{C}^\times)$. It is well-known that if G is a finite abelian group then there is a (non-canonical) isomorphism $G \simeq \mathrm{Hom}(G, \mathbb{C}^\times)$. In particular, any nontrivial multiplicative character generates $\mathrm{Hom}(\mathbb{F}_q^\times, \mathbb{C}^\times)$ since $\mathbb{F}_q^\times$ is cyclic. On the other hand, if ψ is a fixed nontrivial additive character, then every additive character has the form $\psi_a : x \mapsto \psi(ax)$ for some $a \in \mathbb{F}_q$. Indeed, the map $a \mapsto \psi_a$ is an isomorphism $\mathbb{F}_q^+ \to \mathrm{Hom}(\mathbb{F}_q^+, \mathbb{C}^\times)$.

Proposition 2.1.2. *If G is a finite group then we have the following orthogonality relations.*

$$\sum_{\psi \in \mathrm{Hom}(G, \mathbb{C}^\times)} \psi(x) = \begin{cases} |G|, & \text{if } x = 1, \\ 0, & \text{otherwise,} \end{cases} \qquad \sum_{x \in G} \psi(x) = \begin{cases} |G|, & \text{if } \psi \text{ is trivial,} \\ 0, & \text{otherwise.} \end{cases}$$

If χ is a multiplicative character, we extend its definition to all $\mathbb{F}_q$ setting $\chi(0) = 1$ if χ is trivial and $\chi(0) = 0$ otherwise.

2.1.2. Kloosterman sums. There are many different exponential sums that have been studied in the literature. One of the most well-known ones is the Gauss sum

$$G(\chi, \psi) = \sum_{x \in \mathbb{F}_q} \chi(x)\psi(x),$$

where χ is a multiplicative character and ψ is an additive character, which was used by C. F. Gauss to prove the quadratic reciprocity law. Another important family of

exponential sums, which are the main object of this chapter, are the so-called KLOOSTERMAN sums. They have the form

$$S(\psi,\varphi) = -\sum_{x\in\mathbb{F}_q^\times} \psi(x)\varphi(x^{-1}),$$

where both ψ and φ are additive characters, and were introduced to study the problem of the representability of a number by positive definite quadratic forms of four variables [15] (see also [13, Chapter 20]). They also appear, for example, in the Fourier expansions of certain modular forms [13, Chapter 16]. More generally, one can define the KLOOSTERMAN-SALIÉ sums

$$S(\chi;\psi,\varphi) = -\sum_{x\in\mathbb{F}_q^\times} \chi(x)\psi(x)\varphi(x^{-1}),$$

which include both the KLOOSTERMAN sums (χ trivial) and GAUSS sums (φ trivial) as particular cases. The goal of this chapter is to prove WEIL's bound $|S(\psi,\varphi)| \leq 2\sqrt{q}$ for nontrivial characters, and as we shall see the arguments we use follow very closely those of Chapter 1 to prove the RIEMANN hypothesis for curves. Indeed, these exponentials sums are closely related to a particular family of curves for which the RIEMANN hypothesis will give the desired estimates. The key idea is to view $S(\psi,\varphi)$ not as a single object but as a family of sums $S_n(\psi,\varphi)$ which are defined over the extensions $\mathbb{F}_{q^n}$ for $n \geq 1$. This is analogous to base changing a curve over $\mathbb{F}_q$ to the extension $\mathbb{F}_{q^n}$. To do this, recall that for each positive integer $n \geq 1$ we have the trace $\mathrm{Tr}_n : \mathbb{F}_{q^n} \to \mathbb{F}_q$ and the norm $N_n : \mathbb{F}_{q^n} \to \mathbb{F}_q$ functions, which are given explicitly by

$$\mathrm{Tr}_n(x) = \sum_{\sigma\in\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)} \sigma(x), \qquad N_n(x) = \prod_{\sigma\in\mathrm{Gal}(\mathbb{F}_{q^n}/\mathbb{F}_q)} \sigma(x).$$

Then for each integer $n \geq 1$ we define the sums

$$S_n(\chi;\psi,\varphi) = -\sum_{x\in\mathbb{F}_{q^n}^\times} \chi(N_n(x))\psi(\mathrm{Tr}_n(x))\varphi(\mathrm{Tr}_n(x^{-1}))$$

and as in the case of curves we arrange them together into a single object, the zeta function

$$\zeta(\chi;\psi,\varphi;s) = \exp\left(\sum_{n=1}^{\infty} S_n(\chi;\psi,\varphi)q^{-ns}/n\right).$$

Our task now is to evaluate explicitly this function, and that is achieved showing that it is equal to some L-series of the global field $\mathbb{F}_p(X)$.

2.1.3. The L-series of the Kloosterman-Salié sum. Define the norm of a polynomial $h \in \mathbb{F}_q[X]$ by $N(h) = q^{-\deg h}$, and suppose η is a multiplicative character on the group of monic and nonvanishing polynomials in $\mathbb{F}_q[X]$. We extend its definition to all $\mathbb{F}_q[X]$ setting $\eta(h) = 0$ if h is not in this group, and we introduce the L-series $L(s,\eta) = \sum_h \eta(h)N(h)^{-s}$, where h runs over all monic polynomials.

LEMMA 2.1.3. *$\zeta(\chi;\psi,\varphi;s) = (L(s,\eta))^{-1}$, where η is the character defined by*

$$\eta(X^d + a_1X^{d-1} + \cdots + a_{d-1}X + a_d) = \chi((-1)^d a_d)\psi(-a_1)\varphi(-a_{d-1}/a_d).$$

PROOF. By the unique factorisation in $\mathbb{F}_q[X]$ the L-series $L(s,\eta)$ admits the EULER product

$$L(s,\eta) = \prod_P \left(1 - \eta(P)N(P)^{-s}\right)^{-1}$$

where P runs over all monic and irreducible polynomials in $\mathbb{F}_q[X]$. Now taking the logarithm we have

$$\log L(s,\eta) = -\sum_P \log(1 - \eta(P)N(P)^{-s}) = \sum_P \sum_{k\geq 1} \eta(P^k)N(P)^{-ks}$$

and collecting the polynomials P of degree d for each $d \geq 1$ we get

$$\log L(s,\eta) = \sum_{d,k\geq 1} \sum_{\deg P = d} \eta(P^k)q^{-dks}/k = \sum_{n\geq 1} B_n q^{-ns}/n, \text{ where } B_n = \sum_{d=\deg P \mid n} d\eta(P^{n/d}).$$

Notice that the character η defined in the lemma is multiplicative, and furthermore it satisfies

$$a_d = (-1)^d N_d(x), \quad a_1 = -\operatorname{Tr}_d(x), \quad a_{d-1}/a_d = -\operatorname{Tr}_d(x^{-1}).$$

Hence, since the trace and the norm are transitive on extensions, and an irreducible polynomial of degree d has exactly d distinct roots in a fixed algebraic closure $\overline{\mathbb{F}}_q$ as $\mathbb{F}_q$ is perfect, it follows that

$$B_n = \sum_{d\mid n} \sum_{\deg P = d} \sum_{P(x)=0} \chi(N_n(x))\psi(\operatorname{Tr}_n(x))\varphi(\operatorname{Tr}_n(x^{-1})).$$

But the roots of irreducible polynomials of degree divisible by n are exactly the elements in $\mathbb{F}_{q^n}$, so $B_n = -S_n(\chi;\psi,\varphi)$ and the lemma is proved. □

THEOREM 2.1.4. *Suppose ψ and φ are nontrivial. Then*

$$\zeta(\chi;\psi,\varphi;s) = (1 - S(\chi;\psi,\varphi)q^{-s} + \chi(-r)\overline{\chi}(s)q^{1-2s})^{-1}$$

where $\psi(a) = \xi(ar)$, $\varphi(b) = \xi(bs)$ and ξ is any nontrivial additive character.

PROOF. By Lemma 2.1.3 we only have to evaluate $L(s,\chi)$. Notice that we can write

$$L(s,\eta)=\sum_{n\geq 0}A_nq^{-ns},\qquad \text{where } A_n=\sum_{\deg h=d}\eta(h).$$

Clearly $A_0=1$ since $h(X)=1$ is the only monic polynomial of degree 0. Now, the monic polynomials of degree 1 are $X-a$ with $a\in\mathbb{F}_q$, so we have

$$A_1=\sum_{a\in\mathbb{F}_q^\times}\chi(a)\psi(a)\varphi(a^{-1})=-S(\chi;\psi,\varphi).$$

Similarly, the monic polynomials of degree 2 have the form X^2-aX+b for some $a,b\in\mathbb{F}_q$, so

$$A_2=\sum_{a\in\mathbb{F}_q}\sum_{b\in\mathbb{F}_q^\times}\chi(b)\psi(a)\varphi(ab^{-1})=(q-1)\delta_\chi+\sum_{a\in\mathbb{F}_q^\times}\chi(a)\psi(a)\sum_{b\in\mathbb{F}_q^\times}\overline{\chi}(b)\varphi(b)$$

where $\delta_\chi=1$ if χ is trivial and $\delta_\chi=0$ otherwise. To simplify this expression further suppose ψ and φ are nontrivial and assume first that χ is trivial. Then

$$A_2=q-1+\sum_{a\in\mathbb{F}_q^\times}\psi(a)\sum_{b\in\mathbb{F}_q^\times}\varphi(b)=q$$

on using the orthogonality relations. If χ is nontrivial, fix some nontrivial additive character ξ and let $r,s\in\mathbb{F}_q$ be such that $\psi(a)=\xi(ar)$ and $\varphi(b)=\xi(bs)$. Then

$$A_2=\sum_{a\in\mathbb{F}_q^\times}\chi(-ar)\xi(-a)\sum_{b\in\mathbb{F}_q^\times}\overline{\chi}(bs)\xi(b)=\chi(-r)\overline{\chi}(s)|G(\chi,\xi)|^2$$

and it is well-known that $|G(\chi,\xi)|^2=q$. Thus, in any case we have $A_2=\chi(-r)\overline{\chi}(s)q$. Finally, for $n\geq 3$ we have

$$\begin{aligned}A_n&=\sum_{a_1,\dots,a_{n-1}\in\mathbb{F}_q}\sum_{a_n\in\mathbb{F}_q^\times}\chi(a_n)\psi(a_1)\varphi(a_{n-1}/a_n)\\&=q^{n-3}\sum_{a_1\in\mathbb{F}_q}\psi(a_1)\sum_{a_{n-1}\in\mathbb{F}_q}\sum_{a_n\in\mathbb{F}_q^\times}\varphi(a_{n-1}/a_n)\chi(a_n)\end{aligned}$$

and it evaluates to zero since ψ is nontrivial. □

The case of the GAUSS sum is not included in Theorem 2.1.4, but when both χ and ψ are nontrivial and φ is trivial we observe that A_2 vanishes since $\sum_{b\in\mathbb{F}_q^\times}\overline{\chi}(b)=0$, and we still have that $A_n=0$ for all $n\geq 3$. Thus $\zeta(\chi;\psi,1;s)=(1+G(\chi,\psi)q^{-s})^{-1}$, and from this it follows that so-called HASSE-DAVENPORT relation,

$$-G(\chi,\psi)=(-G(\chi,\psi))^n.$$

On the other hand, writing $T = q^{-s}$, we see that the zeta function of the KLOOSTERMAN sum $S = S(\chi, \psi)$ is $(1 - ST + T^2)^{-1}$. Therefore, if α and β are the so-called roots of S, this is, the algebraic integers such that $1 - ST + T^2 = (1 - \alpha T)(1 - \beta T)$, then $\alpha\beta = q$ and we have the relation

$$S_n(\psi, \varphi) = \alpha^n + \beta^n$$

for all $n \geq 1$. In particular the proof of WEIL's bound is reduced to showing that $|\alpha| \leq \sqrt{q}$ and $|\beta| \leq q$. These inequalities, combined with the fact that $\alpha\beta = q$, give $|\alpha| = |\beta| = \sqrt{q}$.

2.1.4. Weil sums. Let $f(X)$ be a polynomial in $\mathbb{F}_q[X]$ of degree $m \geq 1$, ψ be an additive character and χ be a multiplicative character. Another important family of exponential sums are the so-called WEIL sums, which have the form

$$S_n(\psi, f) = -\sum_{x \in \mathbb{F}_{q^n}} \psi(\mathrm{Tr}_n(f(x))), \qquad S_n(\chi, f) = -\sum_{x \in \mathbb{F}_{q^n}} \chi(N_n(f(x))).$$

Let us start studying $S_n(\psi, f)$. As the KLOOSTERMAN sums, these sums have an associated zeta function, namely

$$\zeta(\psi; f; s) = \exp\left(\sum_{n=1}^{\infty} S_n(\psi; f) q^{-ns}/n\right),$$

and a result analogous to Theorem 2.1.4 holds. Precisely, we have the following rationality theorem.

THEOREM 2.1.5. *Suppose $f(X)$ has degree $m \geq 1$ relatively prime to q and let ψ be a nontrivial additive character. Then there exists a polynomial $Q(T)$ of degree at most $m - 1$ such that $Q(0) = 1$ and $\zeta(\psi; f; s) = 1/Q(q^{-s})$.*

In other words, if $\omega_1, \ldots, \omega_{m-1}$ are the roots of $S_n(\psi, f)$, this is, the algebraic numbers such that $Q(T) = \prod_{i=1}^{m-1}(1 - \omega_i T)$, then for all $n \geq 1$ we have

$$S_n(\psi, f) = \omega_1^n + \cdots + \omega_{m-1}^n.$$

The proof of this theorem will follow very closely that of Theorem 2.1.4. First we prove that the zeta function is the inverse of some L-function, and then we show that this L-function is a polynomial of degree at most $m - 1$.

PROOF OF THEOREM 2.1.5. Write $f(X) = \sum_{i=0}^{m} b_i X^i$. To construct the required L-function let η be the multiplicative character on the group of all monic polynomials

$h(X) \in \mathbb{F}_q[X]$ defined by

$$\eta(h) = \chi(f(\alpha_1) + \cdots + f(\alpha_n))$$

where $\alpha_1, \ldots, \alpha_n$ are the roots of h in a fixed algebraic closure $\overline{\mathbb{F}}_q$. Notice that η is well-defined, i.e., $f(\alpha_1) + \cdots + f(\alpha_n) \in \mathbb{F}_q$, since

$$f(\alpha_1) + \cdots + f(\alpha_n) = \sum_{k=0}^{m} b_k s_k(\alpha_1, \ldots, \alpha_n),$$

where $s_k(X_1, \ldots, X_n) = X_1^k + \cdots + X_n^k \in \mathbb{F}_q[X_1, \ldots, X_n]$ is a symmetric polynomial for each $k \geq 0$. Furthermore, we see that $s_k(\alpha_1, \ldots, \alpha_n)$ is exactly $\mathrm{Tr}_n(\alpha_1)$. Now arguing as in Section 2.1.3 we have that

$$L(s, \eta) = \sum_h \eta(h) N(h)^{-s} = \exp\left(\sum_{n \geq 1} B_n q^{-ns}/n \right)$$

where

$$B_n = \sum_{d = \deg P | n} d\eta(P^{n/d}) = \sum_{d|n} \sum_{\deg P = d} \sum_{P(x)=0} \psi(\mathrm{Tr}_n(f(x))) = -S(f; \psi),$$

whence $\zeta(\psi; f; s) = (L(s, \eta))^{-1}$, as claimed.

The last step to prove Theorem 2.1.5 is to show that $A_n = 0$ for all $n \geq m$, where $L(s, \eta) = \sum_{n \geq 0} A_n q^{-ns}$, which implies that $L(s, \eta)$ is a polynomial in q^{-s} of degree at most $m - 1$. We need the following explicit formula, known as Waring's formula, which expresses the symmetric polynomial $s_k(X_1, \ldots, X_n)$ in terms of the elementary symmetric polynomials $\sigma_r(X_1, \ldots, X_n)$.

Lemma 2.1.6. *For all $k \geq 1$ we have*

$$s_k = \sum_{\substack{k = t_1 + 2t_2 + \cdots + nt_n \\ t = t_1 + \cdots + t_n}} (-1)^{k-t} \tfrac{k}{t} C^t_{t_1, \ldots, t_r} \sigma_1^{t_1} \cdots \sigma_n^{t_n}$$

where the sum is over all the tuples $(t_1, \ldots, t_n)$ of non-negative integers satisfying the condition $k = t_1 + 2t_2 + \cdots + nt_n$.

Proof. The formula follows from the following formal computation. We have

$$\begin{aligned} \sum_{k \geq 1} s_k T^k/k &= \sum_{i=1}^{n} \sum_{k \geq 1} s_k X_i^k T^k/k = -\sum_{i=1}^{n} \log(1 - X_i T) \\ &= -\log\Big(\prod_{i=1}^{n} (1 - X_i T) \Big) = -\log\Big(1 - \sum_{i=1}^{n} (-1)^{i-1} \sigma_i T^i \Big) \end{aligned}$$

and using the power series expansion of the logarithm we have that the right hand side is exactly

$$\sum_{t\geq 1}\Big(\sum_{i=1}^{n}(-1)^{i-1}\sigma_i T^i\Big)^t/t = \sum_{k\geq 1} T^k/k \sum_{\substack{k=t_1+2t_2+\cdots+nt_n\\ t=t_1+\cdots+t_n}} (-1)^{k-t} C^t_{t_1,\ldots,t_r}\tfrac{k}{t}\sigma_1^{t_1}\cdots\sigma_n^{t_n}.$$

Now it suffices to compare the coefficients of T^k. □

It now follows from Lemma 2.1.6 that σ_m only appears in the formula for s_k if $m \leq \max(k,n)$, and if $k = m \leq n$ then the coefficient of σ_m is exactly $(-1)^{m-1}m$. Hence, if $n \geq m$, then there exists some polynomial G_n of $m-1$ variables such that

$$f(X_1)+\cdots+f(X_r) = \sum_{k=0}^{m} b_k s_k(X_1,\ldots,X_n) = (-1)^{m-1}mb_m\sigma_m + G_n(\sigma_1,\ldots,\sigma_{m-1})$$

and consequently for the polynomial $h(X) = X^n - a_1X^{n-1} + \cdots + (-1)^n a_n$ we have

$$\eta(h) = \chi((-1)^{m-1}ma_mb_m + G_n(a_1,\ldots,a_{m-1})).$$

Thus, for all $n \geq m$ it follows that

$$\begin{aligned} A_n &= \sum_{a_1,\ldots,a_n\in\mathbb{F}_q} \eta(X^n - a_1X^{n-1} + \cdots + (-1)^n a_n)\\ &= \sum_{a_1,\ldots,a_n\in\mathbb{F}_q} \chi((-1)^{m-1}ma_mb_m + G_n(a_1,\ldots,a_{m-1}))\\ &= q^{n-m}\sum_{a_m\in\mathbb{F}_q}\chi((-1)^{m-1}ma_mb_m)\sum_{a_1,\ldots,a_{m-1}\in\mathbb{F}_q}\chi(G_n(a_1,\ldots,a_{m-1})) \end{aligned}$$

and the first sum evaluates to zero since $(-1)^{m-1}mb_m \neq 0$ as m and q are relatively prime by hypothesis. This finishes the proof of Theorem 2.1.5. □

Now we turn our attention to the other class of Weil sums, namely $S_n(\chi, f)$ for multiplicative characters χ. We define similarly the associated zeta function $\zeta(\chi; f; s)$, and in this case the rationality theorem takes the following form.

Theorem 2.1.7. *Let χ be a multiplicative character of order $r > 1$ and let $f(X)$ be a non-constant monic polynomial which is not an rth power of a polynomial. Let d be the number of distinct roots of f in a fixed algebraic closure of $\mathbb{F}_q$. Then there exists a polynomial $Q(T)$ of degree at most $d-1$ such that $Q(0) = 1$ and $\zeta(\chi; f; s) = 1/Q(q^{-s})$.*

Proof. Again we proceed as in the proof of Theorem 2.1.4. Consider the L-function $L(s,\eta)$, where η is the multiplicative character defined by

$$\eta(h) = \chi(f(\alpha_1)\cdots f(\alpha_n)) = \chi(N_n(f(\alpha_1)))$$

and $\alpha_1, \dots, \alpha_n$ are the roots of $h(X) \in \mathbb{F}_q[X]$ in $\overline{\mathbb{F}}_q$. Then we also have in this case that $\zeta(\chi; f; s) = (L(s, \eta))^{-1}$, and we only need to show that $A_n = 0$ for all $n \geq d$, where $L(s, \eta) = \sum_{n \geq 0} A_n q^{-ns}$. Factor f in $\mathbb{F}_q[X]$ as a product of monic and pairwise distinct irreducible polynomials,

$$f = f_1^{e_1} \cdots f_\ell^{e_\ell},$$

and fix a root β_i of each f_i. Also let $\mathrm{rad}(f) = f_1 \cdots f_\ell$ and put $E_i = \mathbb{F}_q[X]/(f_i) = \mathbb{F}_q(\beta_i)$, which is a finite extension of $\mathbb{F}_q$. Notice that, if h is a monic polynomial of degree n and α is a fixed root of h, then

$$\eta(h) = \chi(N_n(f(\alpha))) = \prod_{i=1}^{\ell} \chi^{e_i}(N_n(f_i(\alpha))) = \chi((-1)^{n \deg f}) \prod_{i=1}^{\ell} \chi^{e_i}(N_{E_i/\mathbb{F}_q}(h(\beta_i))),$$

and if we extend η to all $\mathbb{F}_q[X]$ letting $\eta(h) = 0$ if h is not monic then this formula shows in particular that $\eta : \mathbb{F}_q[X] \to \mathbb{C}^\times$ factors through $\mathbb{F}_q[X]/(\mathrm{rad}(f)) \to \mathbb{C}^\times$ since $\mathrm{rad}(f)(\beta_i) = 0$ for all β_i. Furthermore, by the Chinese Remainder Theorem we have an isomorphism

$$\mathbb{F}_q[X]/(\mathrm{rad}(f)) \to \prod_{i=1}^{\ell} E_i,\ h + \mathrm{rad}(f) \mapsto (h(\beta_1), \dots, h(\beta_\ell)).$$

Let S_n be the set of monic polynomials of degree n. Then the composition

$$S_n \hookrightarrow \mathbb{F}_q[X] \twoheadrightarrow \mathbb{F}_q[X]/(\mathrm{rad}(f)) \simeq \prod_{i=1}^{\ell} E_i,\ h \mapsto (h(\beta_1), \dots, h(\beta_\ell))$$

is surjective if $n \geq d = \deg \mathrm{rad}(f)$, and two polynomials h and h' in S_n have the same image if and only if $h' = h + g\,\mathrm{rad}(f)$ for some polynomial g of degree $< n - d$. Since there are q^{n-d} choices of g we have, for $n \geq d$,

$$A_n = \sum_{h \in S_n} \eta(h) = q^{n-d} \chi((-1)^{n \deg f}) \sum_{x_1 \in E_1} \chi^{e_1}(N_{E_1/\mathbb{F}_q}(x_1)) \cdots \sum_{x_\ell \in E_\ell} \chi^{e_\ell}(N_{E_\ell/\mathbb{F}_q}(x_\ell)),$$

and since not all e_i are multiple of r as f is not an rth power, at least one of $\chi^{e_1}, \dots, \chi^{e_\ell}$ is nontrivial, whence $A_n = 0$ and Theorem 2.1.7 is proved. □

2.2. Weil's bound

2.2.1. Weil's bound for Kloosterman sums. The goal of this section is to prove the following celebrated theorem by Weil which gives the optimal bound for the Kloosterman sums.

THEOREM 2.2.1. *Suppose q is odd and let ψ and φ be nontrivial additive characters on $\mathbb{F}_q$. Then the roots of the Kloosterman sum $S(\psi,\varphi)$ have norm $\sqrt{q}$ and*

$$|S(\psi,\varphi)| \leq 2\sqrt{q}.$$

Since any additive character has the form $x \mapsto \psi(ax)$ for some a, where ψ is a fixed nontrivial character, to prove Theorem 2.2.1 it suffices to consider the sum

$$S(\psi;a,b) = -\sum_{x\in\mathbb{F}_q^\times} \psi(ax+bx^{-1})$$

for $a,b \in \mathbb{F}_q^\times$. The key idea in the proof is to show that the average value of $S(\psi;a,b)$ over all characters ψ is related to the number of rational points on a hyperelliptic curve of the form $Y^2 = f(X)$, where $f(X)$ is some polynomial. The same is true for $S_n(\psi;a,b)$ for all $n \geq 1$, and thus we can argue as in Chapter 1 to conclude that the roots of $S(\psi;a,b)$ have norm $\sqrt{q}$. In any case, we will prove again the RIEMANN hypothesis in this special case in Section 2.3 following STEPANOV's original argument, which is similar to BOMBIERI's but more explicit.

PROPOSITION 2.2.2. *Let α_ψ and β_ψ be the roots of $S(\psi;a,b)$ for each nontrivial ψ, so that $S_n(\psi;a,b) = \alpha_\psi^n + \beta_\psi^n$ for $n \geq 1$. Then*

$$N_n = q^n - 1 - \sum_{\psi\neq 1} (\alpha_\psi^n + \beta_\psi^n)$$

where N_n is the number of $\mathbb{F}_{q^n}$-rational points lying on the curve $Z^2 = (Y^q - Y)^2 - 4ab$.

The proof of the proposition is based on the following preliminary result.

LEMMA 2.2.3. *For all $x \in \mathbb{F}_{q^n}$ we have*

$$\sum_{\psi} \psi(\mathrm{Tr}_n(x)) = |\{y \in \mathbb{F}_{q^n} : y^q - y = x\}|$$

where ψ runs over all additive characters on $\mathbb{F}_q$.

PROOF. First we note that the ARTIN–SCHREIER polynomial $Y^q - Y - x$ splits completely in $\mathbb{F}_{q^n}[Y]$ when $x = y^q - y$ for some $y \in \mathbb{F}_{q^n}$, and it has no roots in $\mathbb{F}_{q^n}$ otherwise. Indeed, if y is a root of $Y^q - Y - x$ in a fixed algebraic closure of $\mathbb{F}_q$, then we see that $y + z$ is also root for all $z \in \mathbb{F}_q$. Hence this polynomial has a root in $\mathbb{F}_{q^n}$ if and only if $y \in \mathbb{F}_{q^n}$, in which case $x = y^q - y$. Conversely, if $x = y^q - y$ then $Y^q - Y - x = f(Y - y)$, where $f(Y) = Y^q - Y$, so $Y^q - Y - x$ has q roots in $\mathbb{F}_{q^n}$. Hence by the additive version of HILBERT's theorem 90 we know that x has the form

$y^q - y$ exactly when $\mathrm{Tr}_n(x) = 0$, and using the orthogonality relations the formula of the lemma follows. □

Proof of Proposition 2.2.2. Put $g(X) = aX + bX^{-1} \in \mathbb{F}_q(X)$. Then by Lemma 2.2.3 we have

$$-\sum_{\psi} S_n(\psi; a, b) = \sum_{\psi} \sum_{x \in \mathbb{F}_{p^n}^{\times}} \psi(\mathrm{Tr}_n(g(x))) = N_n$$

where N_n is the number of points $(x, y) \in \mathbb{F}_{q^n}^{\times} \times \mathbb{F}_{q^n}$ such that $y^q - y = g(x)$. Thus, since $-S_n(1; a, b) = |\mathbb{F}_{q^n}^{\times}| = q^n - 1$, we have

$$N_n = q^n - 1 - \sum_{\psi \neq 1} (\alpha_{\psi}^n + \beta_{\psi}^n)$$

and it suffices to show that N_n is the number of points on the hyperelliptic curve $Z^2 = (Y^q - Y)^2 - 4ab$. But N_n is exactly the number of points $(x, y) \in \mathbb{F}_{q^n} \times \mathbb{F}_{q^n}$ satisfying

$$ax^2 - (y^q - y)x + b = 0,$$

and this is a quadratic equation on x. Since the characteristic of $\mathbb{F}_q$ is odd, the number of solutions to this equation coincides with the number of solutions to the discriminant equation $(y^q - y)^2 - 4ab = z^2$, and this proves the claim. □

Thus, if $C/\mathbb{F}_q$ is the affine algebraic curve defined by the equation $Z^2 - f(Y) = 0$, where $f(Y) = (Y^q - Y)^2 - 4ab$, we have that $N_n = |C(\mathbb{F}_{q^n})|$. Let $\overline{\mathbb{F}}_q$ be a fixed algebraic closure of $\mathbb{F}_q$. Since $4ab \neq 0$, the polynomial $f(Y)$ is not a square in $\overline{\mathbb{F}}_q[Y]$, so $Z^2 - f(Y)$ is irreducible over $\overline{\mathbb{F}}_q[Y, Z]$ and this means that C is geometrically irreducible. Thus by Theorem 1.4.3 we know that for n large enough the estimate $N_n - q = O(q^{\frac{n}{2}})$ holds. In fact, for this family of curves Stepanov's method gives the following explicit result, which we will prove in the following section.

Theorem 2.2.4. *Let $f \in \mathbb{F}_q[X]$ be a polynomial of degree $m = \deg f \geq 3$ and such that $Y^2 - f(X)$ is irreducible over $\overline{\mathbb{F}}_q[X, Y]$. If $q \geq 8m$ and N is the number of pairs $(x, y) \in \mathbb{F}_q \times \mathbb{F}_q$ such that $y^2 = f(x)$, then*

$$|N - q| < 5m\sqrt{q}.$$

Proof of Theorem 2.2.1. By Theorem 2.2.4 it follows that $|N_n - q^n| < 5mq^{\frac{n}{2}}$ if $q^n > 8m$, where $m = \deg f = 2q$, and therefore

$$\Big|\sum_{\psi \neq 1} (\alpha_{\psi}^n + \beta_{\psi}^n)\Big| \leq 1 + 10q^{1+\frac{n}{2}} = O(q^{\frac{n}{2}})$$

for all n such that $q^n \geq 16q$. Then an application of Lemma 1.4.1 yields $|\alpha_\psi| \leq \sqrt{q}$ and $|\beta_\psi| \leq \sqrt{q}$ for all nontrivial ψ, and as we remarked at the end of Section 2.1.3 this is enough to prove Theorem 2.2.1. □

To finish this section we prove that the constant 2 in Theorem 2.2.1 is optimal for fixed a, b and p.

PROPOSITION 2.2.5. *Let a, b and q be fixed. Then $|S(\psi,\varphi)| < 2\sqrt{q}$, and for all $\varepsilon > 0$ there exists some integer $n \geq 1$ such that $|S_n(\psi;a,b)| > (2-\varepsilon)q^{n/2}$.*

PROOF. Write $\alpha_\psi = \sqrt{q}e^{2\pi i\theta}$ for some θ, so that $\beta_\psi = \sqrt{q}e^{-2\pi i\theta}$. Then

$$|S_n(\psi;a,b)| = 2q^{\frac{n}{2}}|\cos(2\pi n\theta)|$$

and either θ is a rational number, in which case $\cos(2\pi n\theta) = 1$ for n large enough, or θ is irrational. In the latter case the sequence $|\cos(2\pi n\theta)|$ is dense in $[0,1]$, so for all $\varepsilon > 0$ there exist infinitely many n such that $|S_n(\psi;a,b)| \geq (2-\varepsilon)q^{n/2}$. This proves the optimality of the bound, and we are left with proving that the bound is never attained, this is, $S(\psi;a,b) = \pm 2\sqrt{q}$ never occurs. Indeed, in this case $S(\psi;a,b)^2$ is multiple of p, the characteristic of $\mathbb{F}_q$. But $S(\psi;a,b)$ is a sum of qth roots of unity, and each of them is congruent to 1 modulo the ideal $(1-\zeta)\mathbb{Z}[\zeta]$, where ζ is a fixed pth root of unity. Hence $S(\psi;a,b) \equiv q-1 \pmod{(1-\zeta)\mathbb{Z}[\zeta]}$, which is a contradiction since $(1-\zeta)\mathbb{Z}[\zeta] \cap \mathbb{Z} = p\mathbb{Z}$. □

2.2.2. Weil's bound for Weil sums. Bounds similar to the one given in Theorem 2.2.1 also hold for the WEIL sums $S(\psi,f)$ and $S(\chi,f)$.

THEOREM 2.2.6. *Suppose $f(X)$ has degree $m \geq 1$ relatively prime to q and let ψ be a nontrivial additive character. Then*

$$|S(\psi,f)| \leq (m-1)\sqrt{q}.$$

THEOREM 2.2.7. *Let χ be a multiplicative character of order $r > 1$ and let $f(X)$ be a non-constant monic polynomial which is not an rth power of a polynomial. Let d be the number of distinct roots of f in a fixed algebraic closure of $\mathbb{F}_q$. Then*

$$|S(\chi,f)| \leq (d-1)\sqrt{q}.$$

As a consequence of Theorems 2.1.5 and 2.1.7, to prove Theorems 2.2.6 and 2.2.7 it suffices to show that the roots of $S(\psi,f)$ and $S(\chi,f)$ have absolute value bounded by

$\sqrt{q}$, which is a weak form of the RIEMANN hypothesis. To handle the sum $S(\chi, f)$, we use the following generalisation of Theorem 2.2.4.

THEOREM 2.2.8. *Let $r \geq 2$ be a divisor of $q-1$ and let $f \in \mathbb{F}_q[X]$ be a polynomial of degree $m \geq 1$ and such that $Y^r - f(X)$ is irreducible over $\overline{\mathbb{F}}_q[X,Y]$. Then for $q \geq 100rm^2$ the number N of $\mathbb{F}_q$-rational points on the curve $Y^r - f(X) = 0$ satisfies*

$$|N - q| < 4mr^{\frac{3}{2}}q^{\frac{1}{2}}.$$

We omit the proof of Theorem 2.2.8 as it is very similar to that of Theorem 2.2.4. We refer the reader to Theorem 6.53 in [16].

Before proving Theorem 2.2.7, we prove the following lemma that gives a criterion to detect when the polynomial $Y^r - f(X)$ is irreducible over $\overline{\mathbb{F}}_q[X,Y]$.

LEMMA 2.2.9. *Let k be a field, $\overline{k}$ be a fixed algebraic closure, and $f \in k[X]$ be a polynomial of degree $m \geq 1$. Let*

$$f(X) = a(X-\alpha_1)^{e_1} \cdots (X-\alpha_d)^{e_d}$$

be the factorisation of f in $\overline{k}[X]$, where $\alpha_1, \ldots, \alpha_d$ are the distinct roots of f. Then $Y^r - f(X)$ is irreducible over $\overline{k}[X,Y]$ if and only if $\gcd(r, e_1, \ldots, e_d) = 1$.

PROOF. Suppose first that $e = \gcd(r, e_1, \ldots, e_d) > 1$, and put

$$g(X) = b(X-\alpha_1)^{e_1/e} \cdots (X-\alpha_d)^{e_d/e}$$

where $b \in \overline{k}$ is any element such that $b^e = a$. If $s = r/e$, then $Y^s - g^e(X)$ divides $Y^r - f(X)$, so $Y^r - f(X)$ is not irreducible over $\overline{k}[X,Y]$. Conversely, suppose that $Y^r - f(X)$ is reducible over $\overline{k}[X,Y]$, and therefore over $K[X,Y]$ for some finite extension K of k, and let $\overline{K(X)}$ be an algebraic closure of $K(X)$. We may assume that K contains a primitive rth roots of unity ζ_r. Then there exists a non-constant polynomial $F(X,Y) \in K[X,Y]$ dividing $Y^m - f(X)$, and if $\xi \in \overline{K(X)}$ is a fixed root of $Y^m - f(X) \in K(X)[Y]$, then we have the factorisation

$$Y^r - f(X) = \prod_{i=1}^{r}(Y - \xi\zeta_r^i),$$

which implies, by unique factorisation of polynomials in $K(X)[Y]$, that there exists some nonempty subset $I \subseteq \{1, \ldots, r\}$ of cardinality $|I| < r$ such that

$$F(X,Y) = \prod_{i \in I}(Y - \xi\zeta_r^i).$$

In particular comparing constant terms it follows that $\xi^{|I|} \in K[X]$. Let $w \leq |I| < r$ be the least positive integer such that $\xi^w \in K(X)$. Clearly if $Y^u \in K(X)$ then u is multiple of w, so $t = r/w > 1$ is an integer and we claim that it divides $\gcd(r, e_1, \ldots, e_d)$. Indeed, write $\xi^w = g/h$ for some $g, h \in K[X]$. Then $fh^t = g^t$, so comparing the multiplicities of the roots we must have that t divides e_i for each i, as wanted. □

Proof of Theorem 2.2.7. Let ℓ be large enough so that $q^\ell \geq 100rm^2$ and $f(X)$ splits completely in $\mathbb{F}_{q^\ell}$, and let

$$f(X) = (X - \alpha_1)^{e_1} \cdots (X - \alpha_d)^{e_d}$$

be the factorisation of f in $\mathbb{F}_{q^\ell}[X]$, where $\alpha_1, \ldots, \alpha_d$ are the distinct roots of $f(X)$. Since f is not an rth power by hypothesis, the number $e = \gcd(m, e_1, \ldots, e_d)$ is a proper divisor of r, so $s = r/e > 1$. Define the polynomial

$$g(X) = (X - \alpha_1)^{e_1/e} \cdots (X - \alpha_d)^{e_d/e} \in \mathbb{F}_{q^\ell}[X]$$

which satisfies $g^e = f$. Then for all $n \geq 1$ we have that

$$S_{\ell n}(\chi, f) = -\sum_{x \in \mathbb{F}_q^{\ell n}} \chi(N_{\ell n}(f(x))) = -\sum_{x \in \mathbb{F}_q^{\ell n}} \chi^e(N_{\ell n}(g(x))) = -\sum_{x \in \mathbb{F}_q^{\ell n}} \lambda(g(x))$$

where λ is the character $\chi^e \circ N_{\ell n}$, which has order $s = r/e$ and in particular it is nontrivial. If we fix a primitive sth root of unity ζ_s, then we can write

$$S_{\ell n}(\chi, f) = -\sum_{i=0}^{s-1} A_i \zeta_s^i, \quad \text{where } A_i = |\{x \in \mathbb{F}_{q^{\ell n}} : \lambda(f(x)) = \zeta_s^i\}|,$$

and if $x_0 \in \mathbb{F}_{q^{\ell n}}$ is any fixed element such that $\lambda(x_0) = \zeta_s$, which exists since λ has order exactly s, then A_i is exactly the number of $x \in \mathbb{F}_{q^{\ell n}}$ satisfying $x_0^{-i} g(x) \in (E^\times)^s$. Furthermore, $A_i = B_i/s$, where B_i is the number of pairs $(x, y) \in \mathbb{F}_{q^{\ell n}} \times \mathbb{F}_{q^{\ell n}}$ such that $x_0^{-i} g(x) = y^s$ and $y \neq 0$. Thus, if N_i is the number of $\mathbb{F}_{q^{\ell n}}$-rational points on the curve $Y^s = x_0^{-i} g(X)$, which is geometrically irreducible by Lemma 2.2.9, then $N_i = q^{\ell n} + O(q^{\frac{\ell n}{2}})$ by Theorem 2.2.8 and $|N_i - B_i|$ is bounded by a constant. Hence it follows that $A_i = q^{\ell n}/s + O(q^{\frac{\ell n}{2}})$, whence

$$S_{\ell n}(\chi, f) = -\sum_{i=0}^{s-1} A_i \zeta_s^i = O(q^{\frac{\ell n}{2}}).$$

If $\omega_1, \ldots, \omega_{d-1}$ are the roots of the sum $S(\chi, f)$, then $S_{\ell n}(\chi, f) = \omega_1^{\ell n} + \cdots + \omega_{d-1}^{\ell n}$ and by Lemma 1.4.1 we deduce that $|\omega_i|^\ell \leq q^{\frac{\ell}{2}}$, which finishes the proof of the theorem. □

To prove Theorem 2.2.7 we need the following result similar to Theorem 2.2.8, which is Theorem 6.59 in [16].

THEOREM 2.2.10. *Let $f \in \mathbb{F}_q[X]$ be a polynomial of degree $m \geq 1$ and suppose* $\gcd(n, q) = 1$. *Let $n \geq 1$ and $b \in \mathbb{F}_{q^n}$. Then, if $N(b)$ is the number of $x \in \mathbb{F}_{q^n}$ such that $\mathrm{Tr}_n(f(x)) = b$, we have*

$$|N(b) - q^{n-1}| < 2m^2 q^{\frac{n}{2}+4}.$$

PROOF OF THEOREM 2.2.7. If $N(b)$ is an in Theorem 2.2.10, then we have

$$S_n(\psi, f) = -\sum_{x \in \mathbb{F}_{q^n}} \psi(\mathrm{Tr}_n(f(x))) = -\sum_{b \in \mathbb{F}_q} N(b)\psi(b) = O(q^{\frac{n}{2}})$$

and now it suffices to argue as in the proof of Theorem 2.2.7 using Lemma 1.4.1. □

2.3. Stepanov's argument

In this section we use STEPANOV's approach to prove Theorem 2.2.4, which gives an estimate for the number of rational points on a distinguished class of hyperelliptic curves. The proof is very similar to the one we gave in Chapter 1, but in this case one constructs a polynomial vanishing at the first coordinates of the points on that curve with high order. The construction is explicit and is based on solving a large system of homogeneous equations which guarantees that the polynomial satisfies the required conditions. Since we are working in positive characteristic derivatives can no longer be used to determine the order of a zero, so we will need to replace them with the so-called HASSE derivatives. They are studied in Section 2.3.1.

We may assume that p is odd since for $p = 2$ the map $x \mapsto x^2$ is an automorphism of $\mathbb{F}_q$, so $N = q$ and the inequality trivially holds. Thus let $p \geq 3$, and define the polynomial $g = f^c$, where $c = \frac{1}{2}(p-1)$. It is well-known that x is a square in $\mathbb{F}_q^\times$ if and only if $x^c = 1$, so

$$N = N_0 + 2N_1, \quad \text{where } N_a = |\{x \in \mathbb{F}_q : g(x) = a\}| \text{ for } a \in \mathbb{F}_q.$$

In particular $N \leq 2(N_0 + N_1)$. On the other hand, since $g(x)$ can only be 0, 1 or -1, we must have $N_{-1} + N_0 + N_1 = q$ and consequently $N \geq 2N_1 = 2q - 2(N_0 + N_{-1})$. Thus to prove Theorem 2.2.4 we only need to show that $N_0 + N_a < \frac{1}{2}q + \frac{5}{2}m\sqrt{q}$ for $a = \pm 1$. But $N_0 + N_a$ is exactly the cardinality of the set

$$S_a = \{x \in \mathbb{F}_q : f(x) = 0 \text{ or } g(x) = a\},$$

and the desired inequality follows easily from the following proposition, whose proof will be given in Section 2.3.2.

PROPOSITION 2.3.1. *Suppose $q \geq 8m$ and let ℓ be an integer such that $\ell \leq q/4 - 1$. Then there exists a nonzero polynomial $r \in \mathbb{F}_q[X]$ of degree*

$$\deg r < mq + \ell\left(\frac{q}{2} + (\ell - 1)m\right)$$

with a zero of order at least ℓ at all points in S_a.

Indeed, let ℓ be the unique integer satisfying $\ell - 1 < \frac{1}{2}\sqrt{q} \leq \ell$. Then

$$\ell \leq \tfrac{1}{2}\sqrt{q} + 1 \leq q/4 - 1 \quad \text{since} \quad q \geq 8m \geq 24 > 16,$$

so by Proposition 2.3.1 we have

$$\ell|S_a| \leq \deg r < mq + \ell\left(\frac{q}{2} + (\ell - 1)m\right) \leq 2m\ell\sqrt{q} + \ell\left(\frac{q}{2} + \frac{m}{2}\sqrt{q}\right),$$

whence $|S_a| < \frac{1}{2}q + \frac{5}{2}m\sqrt{q}$, as claimed.

2.3.1. Hasse derivatives. In characteristic zero we know that the order of vanishing of a polynomial $f(X)$ at some point a is exactly the order of the first nonvanishing derivative of $f(X)$ at a. This, however, is no longer true in positive characteristic as the polynomial X^p shows. The HASSE derivatives are introduced to deal with this problem. They are defined by

$$E^k X^n = \binom{n}{k} X^{n-k}$$

for $n, k \geq 0$, and extended to all polynomials by linearity. In the next two lemmas we study some elementary properties they satisfy which will be needed in Section 2.3.2.

LEMMA 2.3.2. *The Hasse derivatives satisfy the following properties.*

1. *$E^k(fg) = \sum_{j=0}^{k} (E^j f)(E^{k-j} g)$ for all polynomials f, g.*
2. *$E^k(X - a)^r = \binom{r}{k}(X - a)^{r-k}$ for all $k, r \geq 0$ and $a \in \mathbb{F}_q$.*
3. *For all $k, r \geq 0$ with $k \leq r$ and all polynomials f, g there exists a polynomial h of degree bounded by $\deg f + k \deg g - k$ such that $E^k(fg^r) = hg^{r-k}$.*
4. *For all $k < q$ and all $h \in \mathbb{F}_q[X, Y]$ the polynomial $r(X) = h(X, X^q)$ satisfies $E^k r(X) = (E_X^k h)(X, X^q)$, where $E_X^k h$ is the Hasse derivative of h with respect to the variable X.*

PROOF. By linearity of the HASSE derivative it suffices to consider $f(X) = X^n$, $g(X) = X^m$ in part (1) and $h(X, Y) = X^n Y^m$ in part (4) for integers $n, m \geq 0$. Then (1) and (2) follow from the combinatorial identities

$$\sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j} = \binom{n+m}{k}, \qquad \sum_{j=k}^{r} \binom{r}{j}\binom{j}{k} X^{k-j}(-a)^{r-j} = \binom{r}{k}(X-a)^{r-k}$$

while (4) follows from

$$E^k r(X) = E^k X^{n+mq} = \sum_{j=0}^{k} (E^{k-j}X^n)(E^j X^{mq}) = \binom{n}{k} X^{n+mq-k} = (E_X^k h)(X, X^q)$$

on using that $\binom{mq}{j} = 0$ for all $0 < j < q$. Finally, by (1) and induction we have

$$E^k(fg^r) = \sum_{j_1+\cdots+j_r=k} (E^{j_1}f g)(E^{j_2}g)\cdots(E^{j_r}g),$$

and since $k \leq r$ there are at least $r-k$ indices among the j_i's which are zero. This implies that each summand is multiple of g^{r-k}, and its degree is bounded by

$$\deg f + r\deg g - (j_1 + \cdots + j_r) = \deg f + r\deg g - k$$

as required. □

LEMMA 2.3.3. *Let $f \in \mathbb{F}_q[X]$ and suppose that $(E^k f)(a) = 0$ for all $k < \ell$. Then f has a zero of order at least ℓ at a.*

PROOF. By Lemma 2.3.2(2) we may suppose $a = 0$. If we write $f(X) = \sum_{i\geq 0} a_i X^i$ then $(E^k f)(X) = \sum_{i\geq k} a_i \binom{i}{k} X^{i-k}$, so $a_k = 0$ for all $k < \ell$, as wanted. □

2.3.2. Construction of the auxiliary polynomial. To end the proof of Theorem 2.2.4 we only need to construct the polynomial $r \in \mathbb{F}_q[X]$ of Proposition 2.3.1. We assume it has the form

$$r(X) = f^\ell(X) \sum_{j=0}^{J-1} (r_j(X) + s_j(X)g(X))X^{jq}$$

for some large integer J and some polynomials r_j and s_j of degrees $\leq c - m$ that we have to determine. There are three conditions we want r to satisfy. First, each $x \in S_a$ has to be a zero of r of multiplicity at least ℓ. Second, r cannot be identically zero, and finally the degree of r must be bounded by the quantity given in Proposition 2.3.1. Notice that if $x \in S_a$ is already a zero of f then clearly it is a zero of order at least ℓ of r. Hence we may suppose $f(x) \neq 0$.

LEMMA 2.3.4. *For all $k \leq \ell < q$ there exist polynomials $r_j^{(k)}$ and $s_j^{(k)}$ of degrees bounded by $c - m + k(m-1)$ and whose coefficients depend linearly on the coefficients of r_j and s_j such that*

$$E^k r(X) = f^{\ell-k}(X) \sum_{j=0}^{J-1} (r_j^{(k)}(X) + s_j^{(k)}(X) g(X)) X^{jq}.$$

In particular, if $x \in S_a$ then

$$E^k r(x) = f^{\ell-k}(x)\sigma^{(k)}(x), \quad \text{where } \sigma^{(k)}(X) = \sum_{j=0}^{J-1} (r_j^{(k)}(X) + a s_j^{(k)}(X)) X^j.$$

PROOF. We can write $r(X) = h(X, X^q)$, where $h \in \mathbb{F}_q[X, Y]$ is the polynomial

$$h(X, Y) = f^\ell(X) \sum_{j=0}^{J-1} (r_j(X) + s_j(X) f^c(X)) Y^j.$$

Hence, by Lemma 2.3.2(4) we have

$$E^k r = \sum_{j=0}^{J-1} (E^k(f^\ell r_j) + E^k(f^{\ell+c} s_j)) X^{jq}$$

and now it suffices to apply Lemma 2.3.2(3) to conclude. □

We can now finish the proof of Proposition 2.3.1, and thus of Theorem 2.2.4. By Lemmas 2.3.3 and 2.3.4 we have that x is a zero of r of order at least ℓ if $\sigma^{(k)}(x) = 0$ for all $k < \ell$. Consider then the homogeneous system of linear equations

$$\sigma^{(k)}(X) = 0, \quad \text{for } 0 \leq k < \ell$$

on the coefficients of r_j and s_j. Since the number of equations is bounded by

$$\sum_{k=0}^{\ell-1} (\deg \sigma^{(k)} + 1) \leq \sum_{k=0}^{\ell-1} (J + c - m + k(m-1)) = \ell(J + c - m) + \tfrac{1}{2}\ell(\ell-1)(m-1)$$

and the number of coefficients is at least $2(c-m)J$, we have that there is a nontrivial solution as long as the integer J satisfies

$$2(c-m)J > \ell(J + c - m) + \tfrac{1}{2}\ell(\ell-1)(m-1).$$

We can rewrite this inequality as

$$(2(c-m) - \ell)\left(J - \frac{\ell c}{q} - \frac{\ell}{2q}\right) > \frac{\ell}{2}\left(1 + (\ell-1)m\right)$$

and since

$$2(c-m) - \ell = q - 1 - 2m - \ell \geq \frac{q}{2} \quad \text{for} \quad \ell \leq q/4 - 1, \quad m \leq q/8$$

it suffices to take any J satisfying

$$J > \frac{\ell c}{q} + \frac{\ell}{2q} + \frac{\ell}{q}(1 + (\ell - 1)m) = \frac{\ell}{q}\left(\frac{q}{2} + 1 + (\ell - 1)m\right).$$

In fact, let J be the smallest of such integers. Then

$$\begin{aligned}\deg r &\leq \ell \deg f + (c - m) + \deg g + (J - 1)q \\ &< \ell m + c + cm + \ell\left(\frac{q}{2} + 1 + (\ell - 1)m\right)\end{aligned}$$

and since $m \geq 3$ and $m \leq q/8$, $\ell \leq q/4$, we have

$$\ell m + c + cm + \ell \leq mq\left(\frac{1}{4} + \frac{1}{6} + \frac{1}{2} + \frac{1}{12}\right) = mq$$

so

$$\deg r < mq + \ell\left(\frac{q}{2} + (\ell - 1)m\right)$$

as wanted. We are left with proving that r is not identically zero, and this will be a consequence of the geometric irreducibility of the hyperelliptic curve $Y^2 = f(X)$. Indeed, suppose that r is identically zero, and after replacing X with $X - x$ for some appropriate $x \in \mathbb{F}_q$ if necessary suppose that $f(0) \neq 0$. This is possible because the degree of f is less than q. We want to show that $r_j = s_j = 0$ for all j. Assume otherwise and let j_0 be the smallest index such that either r_{j_0} or s_{j_0} is nonzero. Then we have

$$h_0(X) + g(X)h_1(X) = 0$$

where

$$h_0(X) = \sum_{j=j_0}^{J-1} r_j(X)X^{(j-j_0)q}, \qquad h_1(X) = \sum_{j=j_0}^{J-1} s_j(X)X^{(j-j_0)q}.$$

Squaring this relation and multiplying by $f(X)$ we obtain

$$h_0^2(X)f(X) = h_1^2(X)f^q(X) = h_1^2(X)f(X^q) \equiv h_1^2(X)f(0) \pmod{X^q}$$

whence $r_{j_0}^2(X)f(X) \equiv s_{j_0}^2(X)f(0) \pmod{X^q}$. But $\deg r_{j_0}^2 f \leq 2(c - m) + m < q$ and $\deg s_{j_0}^2 \leq 2(c - m) < q$, so we must have $r_{j_0}^2 f = s_{j_0}^2 f(0)$. Since $f(0)$ is a square in $\overline{\mathbb{F}}_q$, so is f in $\overline{\mathbb{F}}_q[X]$, and this contradicts that $Y^2 - f(X)$ is irreducible in $\overline{\mathbb{F}}_q[X, Y]$ by hypothesis.

CHAPTER 3

Heilbronn's exponential sum

3.1. Introduction

Let p be a prime, and define the function $e(x) = \exp(2\pi i x)$, which has period 1. In the 1940s HEILBRONN asked whether the exponential sum

$$H(a) = \sum_{n=1}^{p} e\Big(\frac{an^p}{p^2}\Big),$$

for a relatively prime to p, is $o(p)$ as $p \to \infty$. Notice that, if $n \equiv m \pmod p$, then $n^p \equiv m^p \pmod{p^2}$, so we can let the variable n range over the elements of the finite field $\mathbb{F}_p$ instead, this is, $H(a)$ is a *complete sum*. This problem was not solved until 1996, when HEATH-BROWN [10] proved the estimate $H(a) \ll p^{11/12}$ uniformly on a using STEPANOV's method in a striking way. This bound was subsequently improved to $H(a) \ll p^{7/8}$ by HEATH-BROWN and KONYAGIN [11] using a refined version of the same method, and more recently to $H(a) \ll p^{59/68} \log^{5/34} p$ by SHKREDOV [24] and to $H(a) \ll p^{5/6} \log^{1/6} p$ by SHKREDOV [25]. In this chapter we follow [11] and [28] to prove that the sum

$$H_p(a, \chi) = \sum_{n=1}^{p} \chi(n) e\Big(\frac{an^p}{p^2}\Big)$$

is $\ll p^{7/8}$ as $p \to \infty$ uniformly in a coprime to p for all multiplicative characters χ on $\mathbb{F}_p$. In fact, we will prove that

$$\sum_{r=1}^{p} |H_p(a + rp, \chi)|^4 \ll p^{7/2} \tag{3.1.1}$$

which is enough to deduce the claim.

THEOREM 3.1.2. *$H_p(a, \chi) \ll p^{7/8}$ uniformly in a and χ.*

In particular using this bound we recover the main result of [28] with $p^{7/8}$ instead of $p^{11/12}$. This corollary was improved later to $\ll (h, p-1)^{11/16} p^{7/8}$ by PUCHTA [21].

COROLLARY 3.1.3. *For all $h \geq 1$ we have*

$$\sum_{n=1}^{p-1} e\Big(\frac{an^{hp}}{p^2}\Big) \ll (h, p-1) p^{7/8}.$$

PROOF OF COROLLARY 3.1.3. Replacing h with $(h, p-1)$ we may suppose that h divides $p-1$. Notice that the number of solutions to $n^h \equiv m \pmod p$ is exactly $\sum_\chi \chi(m)$, where χ runs over all multiplicative characters on $\mathbb{F}_p$ satisfying $\chi^h = 1$. Hence

$$\sum_{n=1}^{p-1} e\Big(\frac{an^{hp}}{p^2}\Big) = \sum_{\chi} \sum_{m=1}^{p-1} \chi(m) e\Big(\frac{am^p}{p^2}\Big) = \sum_{\chi} H_p(a,\chi)$$

and since there are h such characters the claim follows from Theorem 3.1.2. □

We start proving the following lemma, which gives a bound for the sum of Equation 3.1.1 in terms of the number of solutions to some polynomial equation. To ease notation write $H(a,\chi) = H_p(a,\chi)$ and define

$$H_0(a,\chi) = \sum_{n=1}^{p-1} \chi(n) e\Big(\frac{an^p}{p^2}\Big) \quad \text{and} \quad H_0(a) = H_0(a,1).$$

LEMMA 3.1.4. *Let* $f(X) = X + \frac{X^2}{2} + \cdots + \frac{X^{p-1}}{p-1} \in \mathbb{F}_p[X]$ *and*

$$M_p = |\{(x_1,x_2) \in \mathbb{F}_p^\times \times \mathbb{F}_p^\times : f(x_1) = f(x_2)\}|.$$

Then $\sum_{r=1}^p |H_0(a+rp,\chi)|^4 \le p^2(p-1+M_p)$.

PROOF. We perform the following simple manipulations. Then from the equality $H_0(a,\chi) = \overline{\chi}(m) H_0(am^p,\chi)$ for all m coprime to p we have

$$(p-1)\sum_{r=1}^{p} |H_0(a+rp,\chi)|^4 = \sum_{r=1}^{p}\sum_{m=1}^{p-1} |H_0((a+rp)m^p,\chi)|^4 \le \sum_{n=1}^{p^2} |H_0(n,\chi)|^4.$$

The last inequality follows since the numbers

$$(a+rp)m^p \quad \text{with} \quad 1 \le r \le p \quad \text{and} \quad 1 \le m \le p-1$$

are pairwise distinct modulo p^2. Indeed, if $(a+rp)m^p \equiv (a+sp)n^p \pmod{p^2}$ then reducing modulo p we find that $m \equiv n \pmod p$ since a is not divisible by p, whence $m = n$ and this readily implies that $r = s$. Thus,

$$(p-1)\sum_{r=1}^{p} |H_0(a+rp,\chi)|^4 \le \sum_{m_1,\dots,m_4=1}^{p-1} \chi(m_1m_2)\overline{\chi}(m_3m_4) \sum_{n=1}^{p^2} e\Big(\frac{(m_1^p + m_2^p - m_3^p - m_4^p)n}{p^2}\Big)$$

and the right hand side is bounded by $p^2 N_p$, where

$$N_p = |\{1 \le m_1, m_2, m_3, m_4 \le p-1 : m_1^p + m_2^p \equiv m_3^p + m_4^p \pmod{p^2}\}|.$$

Reducing modulo p in the above condition we must have $m_1 + m_2 \equiv m_3 + m_4 \pmod p$. Let $b \equiv m_1 - m_3 \pmod p$, so that also $b \equiv m_4 - m_2 \pmod p$. If b is divisible by p then $m_1 = m_3$ and $m_2 = m_4$, so the congruence trivially holds and there are $(p-1)^2$

solutions to it. If b is prime to p write $m_1 \equiv v_1 b \pmod p$ and $m_4 \equiv v_2 b \pmod p$ for some $2 \le v_1, v_2 \le p-1$. Then the congruence reads $v_1^p - (v_1-1)^p \equiv v_2^p - (v_2-1)^p \pmod{p^2}$ and in particular it does not depend on b, whence $N_p = (p-1)^2 + (p-1)M_p'$ where

$$M_p' = |\{2 \le v_1, v_2 \le p-1 : v_1^p - (v_1-1)^p \equiv v_2^p - (v_2-1)^p \pmod{p^2}\}|.$$

But, if p is odd, the fact that $\binom{p}{\ell} \equiv (-1)^{\ell-1} p/\ell \pmod{p^2}$ for all $1 \le \ell \le p-1$ implies

$$v^p - (v-1)^p = 1 - \sum_{\ell=1}^{p-1} \binom{p}{\ell} (-1)^{\ell-1} v^\ell \equiv 1 - p \sum_{\ell=1}^{p-1} \frac{v^\ell}{\ell} \pmod{p^2}$$

so we have $M_p' \le M_p$ and the desired estimate follows. □

For each $u \in \mathbb{F}_p$ define $F(u) = \{x \in \mathbb{F}_p : f(x) = u\}$, so that $M_p = \sum_{u \in \mathbb{F}_p} |F(u)|^2$ and $p = \sum_{u \in \mathbb{F}_p} |F(u)|$. Heath-Brown [10] proved that $|F(u)| \ll p^{2/3}$ uniformly in u, and from this we get the estimate

$$M_p \ll p^{2/3} \sum_{u \in \mathbb{F}_p} |F(u)| = p^{5/2} \tag{3.1.5}$$

which yields $\sum_{r=1}^{p} |H_0(a+rp, \chi)|^4 \ll p^{11/3}$. We remark here that $|F(u)| \ll p^{2/3}$ was first proved by Mit'kin [19], who also proved an analogous result for the truncation of the exponential function. However, the following generalisation of this bound, which is Lemma 7 in [11], allows us to improve the exponent of p in Equation (3.1.5) from $5/2$ to $3/2$. This implies Equation (3.1.1) at once.

Lemma 3.1.6. *For $u \in \mathbb{F}_p$ define $F(u) = \{x \in \mathbb{F}_p : f(x) = u\}$. Then for any subset $U \subseteq \mathbb{F}_p$ we have*

$$|F(U)| = \sum_{u \in U} |F(u)| \ll (p|U|)^{2/3} \quad \textit{where} \quad F(U) = \bigcup_{u \in U} F(u).$$

Corollary 3.1.7. $M_p \ll p^{3/2}$.

Proof of Corollary 3.1.7. Arrange the elements $u_1, \dots, u_p$ of $\mathbb{F}_p$ so that

$$|F(u_1)| \ge |F(u_2)| \ge \cdots \ge |F(u_p)|$$

and consider, for each $T \ge 1$, the set $U = \{u_i : i \le T\}$ of cardinality T. By Lemma 3.1.6 we have $T|F(u_T)| \le |F(U)| \ll (pT)^{2/3}$, so $|F(u_T)| \ll p^{2/3} T^{-1/3}$ and therefore

$$\sum_{N/2 < T \le N} |F(u_T)|^2 \ll p^{2/3} N^{-1/3} \sum_{N/2 < T \le N} |F(u_T)| \ll \min(p^{4/3} N^{1/3}, p^{5/3} N^{-1/3}).$$

Hence, letting $N = 2^k$ with $k \geq 0$, and considering the cases $N \leq p^{1/2}$ and $N > p^{1/2}$ separately, we have

$$\begin{aligned} M_p = \sum_{N=2^k \geq 1} \sum_{N/2 < T \leq N} |F(u_T)|^2 &\ll \sum_{N=2^k \leq p^{1/2}} p^{4/3} N^{1/3} + \sum_{N=2^k > p^{1/2}} p^{5/3} N^{-1/3} \\ &\ll p^{4/3} \sum_{k \leq \log_2 p^{1/2}} 2^{k/3} + p^{5/3} \sum_{k > \log_2 p^{1/2}} 2^{-k/3} \\ &\ll p^{4/3} p^{1/6} + p^{5/3} p^{-1/6} \ll p^{3/2} \end{aligned}$$

as wanted. □

To complete the proof of Theorem 3.1.2 we only need to prove Lemma 3.1.6, and to that end one constructs an auxiliary polynomial $\Psi(X)$ with zeros of large order D at all points in $F(U)$. Thus one gets $|F(U)| \leq \deg \Psi / D$, and for appropriate $\Psi(X)$ and D the lemma will follow. The construction of $\Psi(X)$ is done in Section 3.2 and is motivated by the fact that $f(X)$ is a truncation of the transcendental function

$$-\log(1-X) = X + \frac{X^2}{2} + \frac{X^3}{3} + \cdots \in \mathbb{Q}[[X]].$$

As usual, we also need to show that this polynomial is not identically zero. This is proved in Section 3.3.

3.2. Construction of the auxiliary polynomial

The proof of Lemma 3.1.6 consists in the following application of Stepanov's method. We construct a polynomial $\Phi(X, Y, Z) \in \mathbb{F}_p[X, Y, Z]$ with

$$\deg_X \Phi < A, \quad \deg_Y \Phi < B, \quad \deg_Z \Phi < C$$

and such that the polynomial $\Psi(X) = \Phi(X, f(X), X^p) \in \mathbb{F}_p[X]$ is not identically zero and has a zero of order at least D at all points $x \in F(U)$. This implies that

$$D|F(U)| \leq \deg \Psi(X) \leq \deg_X \Phi + (p-1) \deg_Y \Phi + p \deg_Z \Phi < A + p(B+C) \tag{3.2.1}$$

and by choosing suitable parameters we will obtain the bound given in the lemma. More precisely, we have the following result.

Theorem 3.2.2. *Suppose A, B, C, D are positive real numbers satisfying*

$$D(A + 2D + C)|U| < ABC, \quad D \leq p, \quad AB \leq p.$$

Then $D|F(U)| < A + p(B + C)$.

The condition $D \leq p$ imposed in the theorem allows us to use ordinary derivatives to detect the order of a zero of a polynomial without needing to use Hasse derivatives. Indeed, one easily checks that if $g(X)$ is a polynomial, then x is a zero of g order at least $D \leq p$ if and only if $g^{(n)}(x) = 0$ for all $0 \leq n \leq D-1$.

Before proving Theorem 3.2.2, we prove the following lemma which says that $f(X)$ satisfies some simple differential equations.

Lemma 3.2.3. *For all $n \geq 1$ there exist some polynomials $q_n(X)$ and $h_n(X)$ of degrees bounded by $n+1$ and $n-1$ respectively such that $f(X)$ satisfies the differential equation*

$$(X(1-X))^n \Big(\frac{d}{dX}\Big)^n f(X) = q_n(X) + (X^p - X)h_n(X).$$

Proof. The proof is straightforward by induction. For $n = 1$ we have

$$X(1-X)\frac{d}{dX}f(X) = 1 - X^p,$$

so we can take $h_1(X) = -1$ and $q_1(X) = 0$. Now, if the result is true for $n-1$, then by the induction hypothesis we have

$$(X(1-X))^n \Big(\frac{d}{dX}\Big)^n f(X) = (X(1-X))^n \frac{d}{dX}\left(\frac{q_{n-1} + (X^p - X)h_{n-1}}{(X(1-X))^{n-1}}\right)$$

and some elementary computations show that this is equal to $q_n(X) + (X^p - X)h_n(X)$, where q_n and h_n are the polynomials

$$\begin{aligned} q_n &= X(1-X)q'_{n-1} - X(1-X)h_{n-1} - (n-1)(1-2X)q_{n-1}, \\ h_n &= X(1-X)h'_{n-1} - (n-1)(1-2X)q_{n-1}. \end{aligned}$$

This is enough to prove Lemma 3.2.3. □

Proof of Theorem 3.2.2. Let Φ and Ψ as at the beginning of the section. We have to impose the conditions

$$\Big(\frac{d}{dX}\Big)^n \Psi(X)\Big|_{X=x} = 0 \quad \text{for all} \quad 0 \leq n < D \quad \text{and} \quad x \in F(U). \tag{3.2.4}$$

In fact, in view of Lemma 3.2.3 it will be convenient to replace them with

$$(x(1-x))^n \Big(\frac{d}{dX}\Big)^n \Psi(X)\Big|_{X=x} = 0 \quad \text{for all} \quad 0 \leq n < D \quad \text{and} \quad x \in F(U), \tag{3.2.5}$$

which are equivalent to (3.2.4) unless $x = 0$ or $x = 1$. But this is enough to prove Lemma 3.1.6 since we can replace $F(U)$ with $F(U) \setminus \{0, 1\}$, so if the lemma is true for

$F(U) \setminus \{0,1\}$ then we have

$$|F(U)| \le |F(U) \setminus \{0,1\}| + 2 \ll (p|U|)^{2/3}.$$

Write

$$\Phi(X,Y,Z) = \sum_{a,b,c} \lambda_{a,b,c} X^a Y^b Z^c, \quad \text{so that} \quad \Psi(X) = \sum_{a,b,c} \lambda_{a,b,c} X^a f^b(X) X^{pc}.$$

Since all the derivatives of X^{pc} are zero we see that

$$(X(1-X))^n \Big(\frac{d}{dX}\Big)^n (X^a f^b(X) X^{pc}) = X^{pc} (X(1-X))^n \Big(\frac{d}{dX}\Big)^n (X^a f^b(X)). \tag{3.2.6}$$

On the other hand, for all $j \ge 0$ we have

$$(X(1-X))^j \Big(\frac{d}{dX}\Big)^j X^a = j! C_j^a X^a (1-X)^j$$

so an application of Leibniz' formula yields

$$(X(1-X))^n \Big(\frac{d}{dX}\Big)^n (X^a f^b(X)) = \sum_{j+k=n} j! C_{j,k}^n C_j^a X^a (1-X)^j (X(1-X))^k \Big(\frac{d}{dX}\Big)^k f^b(X)$$

where

$$(X(1-X))^k \Big(\frac{d}{dX}\Big)^k f^b(X) = \sum_{\alpha_1+\cdots+\alpha_b=k} C_{\alpha_1,\dots,\alpha_b}^k \prod_{i=1}^{b} (X(1-X))^{\alpha_i} f^{(\alpha_i)}(X).$$

Using now Lemma 3.2.3 we deduce that the left hand side of Equation (3.2.6) is a linear combination of products of polynomials of the form

$$X^{c+a}(1-X)^{n-(\alpha_1+\cdots+\alpha_\ell)} q_{\alpha_1}(X) \cdots q_{\alpha_\ell}(X) f(X)^{b-\ell} \pmod{X^p - X}$$

with $0 \le \ell \le \min(b,n)$ and $\deg q_{\alpha_i} \le \alpha_i + 1$ for all $1 \le i \le \ell$, and consequently we can write

$$(X(1-X))^n \Big(\frac{d}{dX}\Big)^n \Psi(X) \equiv \sum_{a,b,c} \lambda_{a,b,c} \sum_{0 \le \beta < B} P_\beta(X,a,b,c,n) f^\beta(X) \pmod{X^p - X}$$

for some polynomial $P_\beta(X,a,b,c,n)$ of degree at most $c+a+n+\ell < A+2n+C$. Now, for all $u \in U$ define the polynomial

$$P(X,u,n) = \sum_{a,b,c} \lambda_{a,b,c} \sum_{0 \le \beta < B} P_\beta(X,a,b,c,n) u^\beta.$$

Then the restrictions (3.2.5) read

$$P(x, f(x), n) = 0 \quad \text{for all} \quad 0 \le n < D \quad \text{and} \quad x \in F(U)$$

and therefore it suffices to impose that $P(X,u,n)$ vanishes identically for all $0 \le n < D$ and all $u \in U$. Since the degree of $P(X,u,n)$ is at most $A+2n+C < A+2D+C$, there

are at most $D(A+2D+C)|U|$ linear homogenous equations on the variables $\lambda_{a,b,c}$. The number of variables is ABC, so there exists a nontrivial solution as long as

$$D(A+2D+C)|U| < ABC \quad \text{and} \quad D \leq p. \tag{3.2.7}$$

Hence by Equation (3.2.1) we get the desired bound if $\Psi(X)$ is not identically zero. But this is guaranteed by the following lemma, which finishes the proof of the theorem. □

LEMMA 3.2.8. *The polynomial $\Psi(X)$ is not identically zero if $AB \leq p$.*

The proof of Lemma 3.2.8 is deferred to Section 3.3, where we prove a more general statement. To finish this section, we deduce Lemma 3.1.6 from Theorem 3.2.2.

PROOF OF LEMMA 3.1.6. We claim that the choices

$$A = [p^{2/3}|U|^{-1/3}], \quad B = C = [p^{1/3}|U|^{1/3}], \quad D = [\tfrac{1}{32}p^{2/3}|U|^{-1/3}]$$

satisfy the hypothesis of Theorem 3.2.2. Notice that Lemma 3.1.6 is trivial if $|U| \geq p^{1/2}$, so we may assume that $|U| \leq p^{1/2}$. Both $D \leq p$ and $AB \leq p$ clearly hold, so it suffices to show, for example, that

$$\tfrac{1}{32}p^{2/3}|U|^{2/3}((1+\tfrac{1}{16})p^{2/3}|U|^{-1/3} + p^{1/3}|U|^{1/3}) < (\tfrac{1}{2}p^{2/3}|U|^{-1/3})(\tfrac{1}{2}p^{1/3}|U|^{1/3})^2$$

for p large enough (since $[x] \geq x/2$ as long as $x \geq 1$). But the left hand side is

$$\tfrac{1}{32}p^{4/3}|U|^{1/3}(1+\tfrac{1}{16}+p^{-1/3}|U|^{2/3}) \leq \tfrac{3}{32}p^{4/3}|U|^{1/3} < \tfrac{1}{8}p^{4/3}|U|^{1/3}$$

so the first condition also holds. Finally, with these parameters we obtain the bound

$$D|F(U)| \leq \frac{A+p(B+C)}{D} \ll \frac{p^{4/3}|U|^{1/3}}{p^{2/3}|U|^{-1/3}} = (p|U|)^{2/3}$$

and this completes the proof of Lemma 3.1.6. □

3.3. Nonvanishing of the auxiliary polynomial

To prove that the polynomial $\Psi(X)$ we considered in Section 3.2 does not vanish identically one uses the following result, which is Lemma 3 in [10]. The idea is that $f(X)$ is close to the transcendental function $-\log(1-X)$, so it cannot satisfy any polynomial equation.

LEMMA 3.3.1. *Let $F(X,Y) \in \mathbb{F}_p[X,Y]$ be a polynomial not identically zero and such that $\deg_X F < A$ and $\deg_Y F < B$. Then X^p does not divide $F(X, f(X))$ if $AB \leq p$.*

Let us prove first the nonvanishing of $\Psi(X)$ from Lemma 3.3.1 when $AB \leq p$.

PROOF OF LEMMA 3.2.8. Write

$$\Phi(X,Y,Z)=\sum_{0\le c<C}F_c(X,Y)Z^c$$

for some polynomials $F_c(X,Y)\in\mathbb{F}_p[X,Y]$, and let c_0 be the smallest integer such that $F_{c_0}(X,Y)$ does not vanish identically. Then

$$\Psi(X)=X^{pc_0}\sum_{c_0\le c<C}F_c(X,f(X))X^{p(c-c_0)}$$

and consequently if $\Psi(X)$ is identically zero then $F_{c_0}(X,f(X))$ is divisible by X^p. But this is impossible by Lemma 3.3.1 if $AB\le p$. □

The proof of Lemma 3.3.1 given in [10] is rather technical and is based on the differential equation of $f(X)$ proved in Lemma 3.2.3. YU [28], however, found a simple proof of the following mild generalisation.

LEMMA 3.3.2. *Let $g_0(X),\dots,g_{B-1}(X)$ be polynomials in $\mathbb{F}_p[X]$ with $g_{B-1}(X)$ not identically zero and such that $\deg g_b(X)\le A_b$ for some integers $A_0,\dots,A_{B-1}$ satisfying*

$$A_0\ge A_1\cdots\ge A_{B-1}\quad\textit{and}\quad D=A_0+\cdots+A_{B-1}\le p-B.$$

Then X^{D+B} does not divide

$$F(X)=\sum_{b=0}^{B-1}g_b(X)f^b(X).$$

PROOF. The proof is by induction on B. If $B=1$ then clearly X^{A_0+1} cannot divide $g_0(X)$, which has degree at most A_0. Now let $B>0$, and assume the result is true for all nonnegative integers $<B$. We use induction on $D=D(F)=A_0+\cdots+A_{B-1}$. If $D=0$ then the g_b's are constant, and since $X=0$ is a simple root of $f(X)$ we have that

$$f^b\not\equiv 0\pmod{X^{b+1}},\quad\text{and}\quad f^b\equiv 0\pmod{X^b}$$

for all $0\le b<B$. If X^B divides $F(X)$ then this readily implies that $g_b=0$ for all b, so $F(X)$ is identically zero. Now let $D>0$ be such that $D+B\le p$, and suppose we have proved the lemma for all integers $<D$. For the sake of contradiction suppose there exists some $F(X)$ divisible by X^{D+B}. Then the derivative $F'(X)$ is divisible by X^{D+B-1}, and since $(X-1)f'(X)=1-X^{p-1}$ we easily deduce that

$$(X-1)F'(X)\equiv G(X)\pmod{X^{p-1}}$$

where

$$G(X) = \sum_{b=1}^{B-1}((X-1)g'_{b-1}(X) + bg_b(X))f^{b-1}(X) + (X-1)g'_{B-1}(X)f^{B-1}(X).$$

Since $D+B \le p$ it follows that X^{D+B-1} divides $G(X)$, and consequently it also divides the difference $H(X) = G(X) - rF(X)$, where $r = \deg g_{B-1}(X) \le A_{B-1}$. But

$$\deg((X-1)g'_{b-1}(X) + bg_b(X) - rg_{b-1}(X)) \le A_{b-1}$$

for all $b \le B_1$ since $A_{b-1} \ge A_b$, and

$$\deg((X-1)g'_{B-1}(X) - rg_{B-1}(X)) < r \le A_{B-1},$$

so we have $D(H) < D$ unless $(X-1)g'_{B-1}(X) - rg_{B-1}(X)$ is identically zero. Since $D(H) < D$ is not possible by induction on D we must have

$$(X-1)g'_{B-1}(X) = rg_{B-1}(X)$$

and this implies that $g_{B-1}(X) = a(X-1)^r$ for some a prime to p. But then by induction on B we have

$$(X-1)g'_{b-1}(X) + bg_b(X) - rg_{b-1}(X) = 0 \quad \text{for } 1 \le b \le B-1$$

and in particular for $b = B-1$ this identity reads

$$rg_{B-2}(X+1) - Xg'_{B-2}(X+1) = a(B-1)X^r.$$

Comparing the coefficients of X^r we see that $a(B-1) \equiv 0 \pmod p$, so $a \equiv 0 \pmod p$ and this contradiction completes the proof. □

3.4. Bounds for incomplete Heilbronn sums

In this section we show how to deduce from Equation (3.1.1) the following bound for incomplete Heilbronn sums, which is the corollary to Theorem 1 in [11].

Proposition 3.4.1. *For all $N \le p$ and $M \ge 1$ we have*

$$\sum_{M<n\le M+N} e\Big(\frac{an^p}{p^2}\Big) \ll p^{5/6}N^{1/4}.$$

In the proof we will need the following elementary estimate.

Lemma 3.4.2. *For all real α and integers M and $N \ge 1$ we have*

$$\Big|\sum_{M<n\le M+N} e(n\alpha)\Big| \le \min(N, \|\alpha\|^{-1})$$

where $\|\alpha\|$ is the distance of α to the nearest integer.

PROOF. Replacing α with $\alpha + M$ we may assume $M = 0$, and by periodicity of the exponential function we also may assume that $0 \leq \alpha \leq 1$. The result is clear when α is an integer, and otherwise we have

$$\Big|\sum_{n\leq N} e(n\alpha)\Big| = \left|\frac{1-e(N\alpha)}{1-e(\alpha)}\right| \leq \frac{2}{|\sin \pi\alpha|}.$$

Since $\sin \pi\alpha$ function is concave for $0 < \alpha < 1$ we have $|\sin \pi\alpha| \geq 2\|\alpha\|$, which is enough to prove the claim. □

PROOF OF PROPOSITION 3.4.1. We start writing

$$\begin{aligned}\sum_{M<n\leq M+N} e\Big(\frac{an^p}{p^2}\Big) &= p^{-1}\sum_{r=1}^{p}\sum_{s=1}^{p} e\Big(\frac{as^p}{p^2}\Big) \sum_{M<n\leq M+N} e\Big(\frac{r(s-n)}{p}\Big)\\ &\ll p^{-1}\sum_{r=1}^{p}\Big|\sum_{M<n\leq M+N} e\Big(\frac{rn}{p}\Big)\Big|\Big|\sum_{s=1}^{p} e\Big(\frac{as^p+rsp}{p^2}\Big)\Big|\\ &= p^{-1}\sum_{r=1}^{p}\Big|\sum_{M<n\leq M+N} e\Big(\frac{rn}{p}\Big)\Big||H(a+rp)|.\end{aligned}$$

Then by Lemma 3.4.2 we have

$$\sum_{M<n\leq M+N} e\Big(\frac{an^p}{p^2}\Big) \ll p^{-1}\sum_{r=1}^{p} \min(N, \|r/p\|^{-1})|H(a+rp)|$$

and now HÖLDER's inequality and Equation (3.1.1) yield

$$\begin{aligned}\sum_{M<n\leq M+N} e\Big(\frac{an^p}{p^2}\Big) &\ll p^{-1}\Big(\sum_{r=1}^{p}\min(N, \|r/p\|^{-1})^{4/3}\Big)^{3/4}\Big(\sum_{r=1}^{p}|H(a+rp)|^4\Big)^{1/4}\\ &\ll p^{-1/8}\Big(\sum_{r=1}^{p}\min(N, \|r/p\|^{-1})^{4/3}\Big)^{3/4}.\end{aligned}$$

Finally, since $\|r/p\| = \|(p-r)/p\| = r/p$ for all $1 \leq r < p/2$, we have

$$\begin{aligned}\sum_{r=1}^{p}\min(N, \|r/p\|^{-1})^{4/3} &= 2\sum_{1\leq r<p/2}\min(N, r/p) + N^{4/3}\\ &= 2\sum_{1\leq r<p/N} N^{4/3} + 2\sum_{p/N\leq r<p/2} p^{4/3}r^{-4/3} + N^{4/3} \ll pN^{1/3}\end{aligned}$$

which finishes the proof of Proposition 3.4.1. □

CHAPTER 4

Multidimensional exponential sums

4.1. Deligne's theorems

In this section we review briefly DELIGNE's theory on multidimensional exponential sums, which generalises the one-dimensional case that we have discussed in the previous chapters. While sums in one variable are closely related to some algebraic curves over finite fields, as Chapters 2 and 3 exemplify, these more general sums are related to higher dimensional algebraic varieties. We follow closely the first two sections of [23].

Let $k = \mathbb{F}_q$ be a finite field of q elements and fix an algebraic closure $\overline{k}$ of k. Also let $k_n = \mathbb{F}_{q^n}$ be the unique extension of k of degree n contained in $\overline{k}$, denote by $\mathrm{Tr}_{k_n/\mathbb{F}_p}$ the trace map $k_n \to \mathbb{F}_p$, and define the additive character

$$\psi_{k_n}(x) = \exp(2\pi i \,\mathrm{Tr}_{k_n/\mathbb{F}_p}(x)/p). \tag{4.1.1}$$

If X is an algebraic variety over k and $f \in \mathcal{O}_X(X)$ is a regular function on X, we consider the exponential sum

$$S = S(X, f) = \sum_{x \in X(k)} \psi_k(f(x)),$$

where $f(x)$ denotes the image of f in the residue field $k(x) = k$ of $x \in X(k)$. More generally, for any positive integer $n \geq 1$ we define the companion sums

$$S_n = S(X \times_k k_n, f) = \sum_{x \in X(k_n)} \psi_{k_n}(f(x)),$$

and we construct the associated zeta function of X,

$$Z(X, f, T) = \exp\Big(\sum_{n \geq 1} S_n T^n/n\Big).$$

As usual, this function is equal to a certain Artin L-function.

LEMMA 4.1.2. *We have* $Z(X, f, T) = \prod_x (1 - \psi_{k(x)}(f(x))T^{\deg x})^{-1}$ *where* x *runs over all closed points of* X.

PROOF. The proof is identical to the one given in Section 2.1.3. Taking the logarithm in the product and expanding we obtain

$$\sum_{\ell\geq 1}\sum_{x}\psi_{k(x)}(f(x))^{\ell}T^{k\deg x}/\ell$$

and now collecting the points of degree $d=\deg x\geq 1$ we see that this expression is equal to

$$\sum_{d\geq 1}\sum_{\ell\geq 1}\sum_{\deg x=d}\psi_{k_d}(f(x))^{\ell}T^{\ell d}/\ell=\sum_{n\geq 1}\frac{T^n}{d}\sum_{d=\deg x|n}d\psi_{k_d}(f(x))^{n/d}$$

and the inner sum is exactly S_n by the properties of the trace map and the fact that there are exactly d embeddings $k_d\to k_n$ (cf. Lemma 1.3.3). □

As for the exponential sums studied in Chapter 2, the zeta function $Z(X,f,T)$ is a rational function in T. This result was proved by DWORK using p-adic analysis [7] and by GROTHENDIECK using ℓ-adic cohomology [9].

THEOREM 4.1.3. *The zeta function $Z(X,f,T)$ of an algebraic variety X is a rational function of T.*

Thus we can write $Z(X,f,T)$ as the quotient $A(T)/B(T)$ of two polynomials, and if we let $\alpha_1,\dots,\alpha_r$, $\beta_1,\dots,\beta_s$ be the algebraic numbers such that $A(T)=\prod_i(1-\alpha_iT)$ and $B(T)=\prod_j(1-\beta_jT)$, then the sum S_n can be written explicitly as

$$S_n=\sum_{i=1}^{r}\alpha_i^n-\sum_{j=1}^{s}\beta_j^n.$$

Furthermore, an analogous version of the RIEMANN hypothesis also holds for S_n. This is the content of the following celebrated theorem by DELIGNE, whose proof crucially uses the LEFSCHETZ fixed-point theorem.

THEOREM 4.1.4 (DELIGNE). *There exist algebraic integers ω_{ij} of absolute value $q^{r(i,j)/2}$ for some integer $r(i,j)$, called the weight of ω_{ij}, such that*

$$S_n=\sum_{i,j}(-1)^i\omega_{ij}^n. \tag{4.1.5}$$

Furthermore, any conjugate of ω_{ij} over $\mathbb{Q}$ has also absolute value $q^{r(i,j)/2}$.

We end the section with the following result from [4], which gives an upper bound for the number of roots of maximal weight in terms of an explicit bound for S_n. This will be needed in the next section.

LEMMA 4.1.6. *If $r \geq 1$ is a positive integer and $C > 0$ is such that $|S_n| \leq Cq^{\frac{nr}{2}}$ for every large n, then the roots of S_n have weight at most r, and there are at most C^2 roots of weight exactly r.*

PROOF. Let $B \geq 0$ be the number of roots of weight r. The result will follow if we prove that for all $\varepsilon > 0$ there exists infinitely many n for which the lower bound

$$|S_n| > (B^{\frac{1}{2}} - \varepsilon)q^{\frac{nr}{2}} \tag{4.1.7}$$

holds. Indeed, in this case we see that $C > B^{\frac{1}{2}} - \varepsilon$ for all $\varepsilon > 0$, so $C \geq B^{\frac{1}{2}}$, as wanted.

To prove that (4.1.7) holds for infinitely many n, notice that we can write

$$S_n q^{-\frac{nr}{2}} = \sum_{\ell=1}^{s} a_\ell z_\ell^n + O(q^{-\frac{n}{2}})$$

for some pairwise distinct complex numbers $z_1, \ldots, z_s$ of absolute value 1 and some nonzero integers $a_1, \ldots, a_s$ such that $\sum_{\ell=1}^{s} |a_\ell| = B$, so

$$\limsup_{n\to\infty} |S_n| q^{-\frac{nr}{2}} = \limsup_{n\to\infty} \Big|\sum_{\ell=1}^{s} a_\ell z_\ell^n\Big|.$$

But we have

$$\sum_{n=0}^{N} \Big|\sum_{\ell=1}^{s} a_\ell z_\ell^n\Big|^2 = (N+1)\sum_{\ell=1}^{s} |a_\ell|^2 + \sum_{\ell\neq m} |a_\ell a_m| \Big|\sum_{\ell=1}^{s} (z_\ell \overline{z}_m)^n\Big| = (N+1)\sum_{\ell=1}^{s} |a_\ell|^2 + O(1)$$

where the last equality follows from the fact that the z_ℓ's are pairwise distinct. From this it readily follows that

$$\limsup_{n\to\infty} |S_n| q^{-\frac{nr}{2}} = \Big(\sum_{\ell=1}^{s} |a_n|^2\Big)^{\frac{1}{2}} \geq B^{\frac{1}{2}}$$

and the proof of the lemma is finished. □

4.2. An example by Birch and Bombieri

Let p be a prime and q be a power of p. Also let ψ_q be the additive character of $\mathbb{F}_q$ defined in Equation (4.1.1). In this section we study the exponential sum

$$S(q) = S(q, \alpha, \beta, \gamma) = \sum_{x,y,z\in\mathbb{F}_p}{}' \psi_q(x^{-1}y + (x+\alpha)^{-1}z + \beta y^{-1} + \gamma z^{-1}) \tag{4.2.1}$$

where α, β and γ are elements of $\mathbb{F}_q^\times$. The prime in the summation sign means that the variables x, y, z are restricted to those values for which the argument of ψ_q is defined, namely they are nonzero and $x + \alpha \neq 0$. The particular case of q prime was needed in the celebrated paper [8] by FRIEDLANDER and IWANIEC, where they apply ideas developed by BURGESS in [6] to obtain bounds for certain averages of KLOOSTERMAN

sums. This in turn leads them to obtain interesting results on the following problem in additive number theory. Let $d_3(m)$ be the number of representations of a positive integer m as product of three positive integers, that is,

$$d_3(m) = \#\{(m_1, m_2, m_3) \in \mathbb{Z}^3_{\geq 1} : m_1 m_2 m_3 = m\},$$

and define the average function

$$D_3(X, q, a) = \sum_{\substack{m \leq X \\ m \equiv a \bmod q}} d_3(m)$$

for positive integers q and a. Then they prove that the asymptotic formula

$$D_3(X, q, a) = \frac{X}{\phi(q)} P(\log X) + O_\varepsilon(X^{A+\varepsilon} q^{-B})$$

holds uniformly in the range $X^{92/185} < q < X^{58/115}$ for any a relatively prime to q, where A and B are some explicit positive numbers and P is a certain quadratic polynomial. Notice that $58/115 = 1/2 + 1/230$.

The main theorem we prove in this section on the sum given in Equation (4.2.1) is the following.

THEOREM 4.2.2. *Let q be odd and $\alpha\beta\gamma \neq 0$. Then there exist absolute constants C_0 and C_1 such that*

$$|S(q, \alpha, \beta, \gamma)| \leq C_1 q^{\frac{3}{2}}, \quad \textit{for all primes } p \geq C_0.$$

In [2] BIRCH and BOMBIERI give two different proofs of Theorem 4.2.2. The first one depends deeply in algebraic geometry while the second one is of a more computation nature. Here we will follow the second one as it is closer to the spirit of the previous chapters.

PROOF OF THEOREM 4.2.2. We start observing that multiplying the variables x, y and z by α in the sum (4.2.1) we may assume that $\alpha = 1$. Hence we only need to study

$$S(\beta, \gamma) = S(q, \beta, \gamma) = \sum_{x,y,z \in \mathbb{F}_p}{}' \psi_q(x^{-1}y + (x+1)^{-1}z + \beta y^{-1} + \gamma z^{-1}).$$

with $\beta, \gamma \neq 0$. In the next lemma we show that this sum can be simplified further in terms of the KLOOSTERMAN sum

$$K(a) = \sum_{x \in \mathbb{F}_q^\times} \psi_q(x^{-1} + ax), \quad a \in \mathbb{F}_q.$$

LEMMA 4.2.3. $S(\beta,\gamma)=\sum_{x\neq 0,-1} K\big(\frac{\beta}{x}\big)K\big(\frac{\gamma}{1+x}\big)$.

PROOF. Making the change of variables $Y=\beta^{-1}y$ and $Z=\gamma^{-1}z$ we have

$$\begin{aligned} S(\beta,\gamma)&=\sum_x\sum_y \psi_q(x^{-1}y+\beta y^{-1})\sum_z \psi_q((x+1)^{-1}z+\gamma z^{-1})\\ &=\sum_x\sum_Y \psi_q(Y^{-1}+x^{-1}\beta Y)\sum_Z \psi_q(Z^{-1}+(x+1)^{-1}\gamma Z^{-1})\\ &=\sum_x K(x^{-1}\beta)K((x+1)^{-1}\gamma)\end{aligned}$$

as wanted. □

LEMMA 4.2.4. *For $b,c\in\mathbb{F}_q^\times$ we have*

$$\sum_{a\in\mathbb{F}_q} K(ab)K(ac)=\delta_{b,c}q^2-q$$

where $\delta_{b,c}=1$ if $b=c$ and $\delta_{b,c}=0$ otherwise.

PROOF. We have

$$\begin{aligned}\sum_{a\in\mathbb{F}_q} K(ab)K(ac)&=\sum_{x,y\in\mathbb{F}_q^\times}\psi_q(x^{-1}+y^{-1})\sum_{a\in\mathbb{F}_q}\psi_q(a(bx+cy))\\ &=q\sum_{x\in\mathbb{F}_q^\times}\psi_q(x^{-1}(1+b^{-1}c))\end{aligned}$$

and the rightmost sum evaluates to $q-1$ if $b=c$ and to -1 if $b\neq c$. This implies the claim. □

From Lemma 4.2.3 we deduce that

$$\begin{aligned}\sum_{\beta\in\mathbb{F}_q}(S(\beta,\gamma))^2&=\sum_{\beta\in\mathbb{F}_q}\sum_{x,y\neq 0,-1}K\Big(\frac{\beta}{x}\Big)K\Big(\frac{\gamma}{1+x}\Big)K\Big(\frac{\beta}{y}\Big)K\Big(\frac{\gamma}{1+y}\Big)\\ &=\sum_{x,y\neq 0,-1}K\Big(\frac{\gamma}{1+x}\Big)K\Big(\frac{\gamma}{1+y}\Big)\sum_{\beta\in\mathbb{F}_q}K\Big(\frac{\beta}{x}\Big)K\Big(\frac{\beta}{y}\Big)\\ &=q^2\sum_{x\neq 0,-1}\Big(K\Big(\frac{\gamma}{1+x}\Big)\Big)^2-q\sum_{x,y\neq 0,-1}K\Big(\frac{\gamma}{1+x}\Big)K\Big(\frac{\gamma}{1+y}\Big)\end{aligned}$$

and now the change of variables $X=(1+x)^{-1}$ and $Y=(1+y)^{-1}$ yields

$$\begin{aligned}\sum_{\beta\in\mathbb{F}_q}(S(\beta,\gamma))^2&=q^2\sum_{X\neq 0,1}(K(X\gamma))^2-q\Big(\sum_{X\neq 0,1}K(X\gamma)\Big)^2\\ &=q^2(q^2-q-1-(K(\gamma))^2)-q(1-K(\gamma))^2\end{aligned}$$

on using Lemma 4.2.4 and that $\sum_{a\in\mathbb{F}_q} K(ab) = 0$ if $b \neq 0$. Thus we obtain

$$\sum_{\beta\in\mathbb{F}_q} (S(\beta,\gamma))^2 = q^4 - q^3 - q^2 - q^2(K(\gamma))^2 - q(1-K(\gamma))^2$$

and in particular we have that $S(\beta,\gamma) \leq q^2 + O(q^{\frac{3}{2}})$ since $|K(\gamma)| \leq 2q^{\frac{1}{2}}$ (Theorem 2.2.1). Hence for any fixed $C > 1$ the inequality $S(\beta,\gamma) \leq Cq^2$ holds for large enough q, so we are under the hypothesis of Lemma 4.1.6 and (using any $C < \sqrt{2}$) it follows that for each fixed $\gamma \neq 0$ the sum $S(\beta,\gamma)$ has roots of weight at most $\frac{3}{2}$ with at most one exception, which must have weight 2. If $S(\beta,\gamma)$ has a root of weight 2 we say that the pair (β,γ) is *exceptional*. If (β,γ) is exceptional, then from Equation (4.1.5) we have that

$$S(\beta,\gamma) = \varepsilon q^2 + O(q^{\frac{3}{2}})$$

where $\varepsilon = \varepsilon(\beta,\gamma)$ is either 1 or -1. Notice that if no exceptional pair exists, then $S(\beta,\gamma) = O(q^{\frac{3}{2}})$ and the theorem follows. Therefore assume that at least one exceptional pair exists. We will see that under this hypothesis the number q must remain bounded by some constant C_0.

LEMMA 4.2.5. *The pair (β,γ) is exceptional with $\varepsilon = \varepsilon(\beta,\gamma)$ if and only if*

$$\sum_{x\neq 0,-1} \Big(K\Big(\frac{\beta}{x}\Big) - \varepsilon K\Big(\frac{\gamma}{1+x}\Big)\Big)^2 = O(q^{\frac{3}{2}}).$$

PROOF. Expanding the sum on the left-hand side we see that it is equal to

$$\sum_{x\neq 0,-1} (K(x^{-1}\beta x))^2 + \sum_{x\neq 0,-1} (K((1+x)^{-1}\gamma x))^2 - 2\varepsilon S(\beta,\gamma)$$

and now using Lemma 4.2.4 this expression simplifies to

$$2(q^2-q-1) - (K(\beta))^2 - (K(\gamma))^2 - 2\varepsilon S(\beta,\gamma) = 2q^2 + O(q) - 2\varepsilon S(\beta,\gamma).$$

Here we have used that $K(a) = O(q^{\frac{1}{2}})$ (Theorem 2.2.1). But this is $O(q^{\frac{3}{2}})$ if and only if $S(\beta,\gamma) = \varepsilon q^2 + O(q^{\frac{3}{2}})$, as claimed. □

As a consequence of Lemma 4.2.5 and the change of variables $x \mapsto -1-x$ we see that if (β,γ) is exceptional, then so is $(-\gamma,-\beta)$ and they have the same ε. We next prove that there is a composition law for exceptional pairs.

LEMMA 4.2.6. *If (β,γ) and (β',γ') are exceptional and $\beta \neq -\gamma'$, then*

$$(\beta'',\gamma'') = (\beta,\gamma) \oplus (\beta',\gamma') = \Big(\frac{\beta\beta'}{\beta+\gamma'}, \frac{\gamma\gamma'}{\beta+\gamma'}\Big)$$

is also exceptional and $\varepsilon(\beta'',\gamma'') = \varepsilon(\beta,\gamma)\varepsilon(\beta',\gamma')$.

PROOF. Put $\varepsilon = \varepsilon(\beta, \gamma)$ and $\varepsilon' = \varepsilon(\beta', \gamma')$. Since the pair $(-\gamma', -\beta')$ is exceptional we have, by Lemma 4.2.5, that

$$\sum_{x \neq 0, -1} \Big(K\Big(\frac{-\gamma'}{x}\Big) - \varepsilon K\Big(\frac{-\beta'}{1+x}\Big)\Big)^2 = O(q^{\frac{3}{2}})$$

and now the change of variables $x \mapsto (-\gamma'/\beta)x$ yields

$$\sum_{x \neq 0, -1} \Big(K\Big(\frac{\beta}{x}\Big) - \varepsilon K\Big(\frac{-\beta\beta'}{\beta - \gamma' x}\Big)\Big)^2 = O(q^{\frac{3}{2}}).$$

But (β, γ) is an exceptional pair, so using Lemma 4.2.5 again and the triangle inequality one easily obtains that

$$\sum_{x \neq 0, -1} \Big(K\Big(\frac{\gamma}{1+x}\Big) - \varepsilon\varepsilon' K\Big(\frac{-\beta\beta'}{\beta - \gamma' x}\Big)\Big)^2 = O(q^{\frac{3}{2}}).$$

Finally, after the change of variables $y = ((\beta/\gamma') + 1)x + (\beta/\gamma')$ this equality reads

$$\sum_{y \neq 0, -1} \Big(K\Big(\frac{\beta\beta'/(\beta+\gamma')}{y}\Big) - \varepsilon\varepsilon' K\Big(\frac{\gamma\gamma'/(\beta+\gamma')}{1+y}\Big)\Big)^2 = O(q^{\frac{3}{2}})$$

and this finishes the proof of the lemma. □

LEMMA 4.2.7. *If (β, γ) is exceptional, so is $(m^2\beta, m^2\gamma)$ for all $m \in \mathbb{F}_p^\times$.*

PROOF. Making the change of variables $x \mapsto mx$ we easily see that

$$K(am^2) = \sum_{x \in \mathbb{F}_q^\times} \psi_q(m(x^{-1} + ax)) = \sigma K(a)$$

where $\sigma \in \mathrm{Gal}(\mathbb{Q}(e^{2\pi i/p})/\mathbb{Q})$ is the automorphism $e^{2\pi i/p} \mapsto e^{2\pi i m/p}$. Therefore from Lemma 4.2.3 and Equation (4.1.5) we deduce that

$$S(m^2\beta, m^2\gamma) = \sigma S(\beta, \gamma) = \sum_{i,j} (-1)^i \sigma(\omega_{ij})^n.$$

But by the second part of Theorem 4.1.4 the weight of $\sigma(\omega_{ij})$ is exactly that of ω_{ij}, so ω_{ij} has weight 2 if and only if $\sigma(\omega_{ij})$ has weight 2. □

LEMMA 4.2.8. *If (β, γ) is exceptional, so is $(m\gamma, m\gamma)$ for each $m \in \mathbb{F}_q^\times$, and furthermore $\varepsilon(m\gamma, m\gamma) = 1$.*

PROOF. By Lemmas 4.2.6 and 4.2.7 the pair

$$(m^2\beta, m^2\gamma) \oplus (-\gamma, -\beta) = \left(\frac{m^2}{1-m^2}\gamma, \frac{m^2}{1-m^2}\gamma \right)$$

is exceptional if $m \neq \pm 1$. Now, since $(\gamma,\gamma) \oplus (\gamma,\gamma) = (\gamma/2, \gamma/2)$ for any γ, composing this exceptional pair with itself n times we have that

$$\left(\frac{m^2}{1-m^2}\frac{\gamma}{n}, \frac{m^2}{1-m^2}\frac{\gamma}{n}\right)$$

is also exceptional for each $n \in \mathbb{F}_p^\times$. In particular this shows that $(m\gamma, m\gamma)$ is exceptional for each $m \in \mathbb{F}_p^\times$.

To prove the second part, we show more generally that if (γ,γ) is exceptional, then $\varepsilon = \varepsilon(\gamma,\gamma)$ is 1. Indeed, the exceptional pair $(\gamma,\gamma) \oplus (\gamma,\gamma) = (\gamma/2,\gamma/2)$ has $\varepsilon(\gamma/2,\gamma/2) = \varepsilon^2 = 1$. Now, if $1 \leq n < p$ is an inverse of 2 modulo p, then composing this pair n times we obtain that the pair $(\gamma/2n, \gamma/2n) = (\gamma,\gamma)$ has also $\varepsilon = 1$, as wanted. □

From Lemma 4.2.8 we have that

$$\sum_{m\in\mathbb{F}_p^\times} S(m\gamma, m\gamma) = (p-1)q^2 + O(q^{\frac{3}{2}}p). \tag{4.2.9}$$

On the other hand, if Tr is the absolute trace $\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}$, then this sum can be written as

$$\begin{aligned}\sum_{m\in\mathbb{F}_p^\times} S(m\gamma, m\gamma) &= \sum_{x,y,z} \psi_q(x^{-1}y + (x+1)^{-1}z) \sum_{m\in\mathbb{F}_p^\times} \psi_q(\gamma m(y^{-1}+z^{-1})) \\ &= p \sum_{\substack{x,y,z \\ \mathrm{Tr}(\gamma(y^{-1}+z^{-1}))=0}} \psi_q(x^{-1}y + (x+1)^{-1}z) - \sum_{x,y,z} \psi_q(x^{-1}y + (x+1)^{-1}z) \\ &= p \sum_{\substack{x,s,u \\ \mathrm{Tr}(\gamma s)=0}} \psi_q\Big(\frac{1}{xu} + \frac{1}{(x+1)(s-u)}\Big) - (q-2)\end{aligned}$$

where $s = y^{-1} + z^{-1}$ and $u = y^{-1}$. In the particular case when $q = p$ is prime the last sum can be evaluated explicitly. Indeed, the trace map Tr is the identity and the condition $\mathrm{Tr}(\gamma s) = 0$ reads $s = 0$, whence the sum is equal to

$$\sum_{x\neq 0,-1} \sum_{u\neq 0} \psi_q\Big(\frac{1}{xu} - \frac{1}{(x+1)u}\Big) = p - 2.$$

Hence

$$\sum_{m\in\mathbb{F}_p^\times} S(m\gamma, m\gamma) = O(p^2)$$

and comparing this bound with Equation (4.2.9) we obtain $(p-1)p^2 + O(p^{\frac{5}{2}}) = O(p^2)$, this is, $p = O(1)$. This proves that there exists a constant $C_0 > 0$ such that an exceptional pair exists only if $p < C_0$, as wanted.

Finally assume that q is arbitrary. We shall use the following result, which is Theorem 5 in [12].

THEOREM 4.2.10. *Let $f(x_1, x_2, x_3)$ and $g(x_1, x_2, x_3)$ be polynomials with coefficients in $\mathbb{F}_q$, and consider the exponential sum*

$$S(f, g) = \sum_{g(x_1,x_2,x_3)=0} \psi_q(f(x_1, x_2, x_3)).$$

Suppose that the variety

$$f(x_1, x_2, x_3) - t = 0, \quad g(x_1, x_2, x_3) = 0$$

is geometrically irreducible for almost all t, and that it is a curve for all t, possibly empty. Then $S(f, g) = O(q)$.

In particular consider the polynomials

$$f_s(x, u, z) = (xs + s - u)z,$$
$$g_s(x, u, z) = xu(x + 1)(s - u)z - 1.$$

Then

$$\frac{1}{xu} + \frac{1}{(x + 1)(s - u)} = \frac{f_s(x, u, z)}{g_s(z, u, z) + 1}$$

so

$$\sum_{x,u} \psi_q\Big(\frac{1}{xu} + \frac{1}{(x + 1)(s - u)}\Big) = \sum_{g(x_1,x_2,x_3)=0} \psi_q(f(x_1, x_2, x_3))$$

for each fixed s, and by Theorem 4.2.10 this sum is $O(q)$. Hence by Equation (4.2.9) we deduce that $pq^2 + O(pq^{\frac{3}{2}}) = O(pq)$, whence $q = O(1)$, as wanted. □

Bibliography

[1] M. F. Atiyah and I. G. MacDonald. *Introduction to commutative algebra*. Addison-Wesley-Longman, 1969.

[2] Bryan J. Birch and Enrico Bombieri. "On some exponential sums". In: *Annals of Mathematics* 121.2 (1985), pp. 345–350.

[3] Enrico Bombieri. "Counting points on curves over finite fields". In: *Séminaire Bourbaki* 15 (1974), pp. 234–241.

[4] Enrico Bombieri. "On exponential sums in finite fields, II". In: *Inventiones mathematicae* 47.1 (1978), pp. 29–39.

[5] Nicolas Bourbaki. *Algebra II: Chapters 4-7*. Springer Science & Business Media, 2013.

[6] David A. Burgess. "On character sums and primitive roots". In: *Proceedings of the London Mathematical Society* 3.1 (1962), pp. 179–192.

[7] Bernard Dwork. "On the zeta function of a hypersurface". In: *Publications Mathématiques de l'IHÉS* 12 (1962), pp. 5–68.

[8] John B. Friedlander and Henryk Iwaniec. "Incomplete Kloosterman sums and a divisor problem". In: *Annals of Mathematics* 121.2 (1985), pp. 319–344.

[9] Alexander Grothendieck. "Formule de Lefschetz et rationalité des fonctions L". In: *Séminaire Bourbaki* 9 (1964), pp. 41–55.

[10] D. R. Heath-Brown. "An estimate for Heilbronn's exponential sum". In: *Analytic number theory, Vol. 2 (Allerton Park, IL, 1995)*. Progr. Math. 139 (1996), pp. 451–463.

[11] D. R. Heath-Brown and S. Konyagin. "New bounds for Gauss sums derived from kth powers, and for Heilbronn's exponential sum". In: *Q. J. Math.* 51.2 (2000), pp. 221–235.

[12] C. Hooley. "On exponential sums and certain of their applications". In: *Number theory days, 1980 (Exeter, 1980)*. Vol. 56. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge-New York, 1982, pp. 92–122.

[13] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory*. Vol. 53. American Mathematical Soc., 2004.

[14] Gerald J. Janusz. *Algebraic number fields*. Vol. 7. American Mathematical Soc., 1996.

[15] H. D. Kloosterman. "On the representation of numbers in the form $ax^2 + by^2 + cz^2 + dt^2$". In: *Acta mathematica* 49.3-4 (1926), pp. 407–464.

[16] Rudolf Lidl and Harald Niederreiter. *Finite Fields*. 2nd ed. Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1996.

[17] Qing Liu. *Algebraic Geometry and Arithmetic Curves*. Oxford graduate texts in mathematics. Oxford University Press, 2002.

[18] Dino Lorenzini. *An invitation to arithmetic geometry*. Vol. 9. American Mathematical Soc., 1996.

[19] D. A. Mit'kin. "Stepanov method of the estimation of the number of roots of some equations". In: *Mathematical Notes* 51.6 (1992), pp. 565–570.

[20] Bjorn Poonen. *Rational points on varieties*. Vol. 186. American Mathematical Soc., 2017.

[21] Jan-Christoph Puchta. "Remark on a paper of Yu on Heilbronn's exponential sum". In: *Journal of Number Theory* 87.2 (2001), pp. 239–241.

[22] Michael Rosen. *Number Theory in Function Fields*. Graduate Texts in Mathematics 210. Springer-Verlag New York, 2002.

[23] Jean-Pierre Serre. "Majorations de sommes exponentielles". In: *Astérisque* 41.2 (1977).

[24] I. D. Shkredov. "On Heilbronn's exponential sum". In: *Q. J. Math.* 64.4 (2013), pp. 1221–1230.

[25] Ilya D. Shkredov. "On exponential sums over multiplicative subgroups of medium size". In: *Finite Fields and Their Applications* 30 (2014), pp. 72–87.

[26] Joseph H. Silverman. *The arithmetic of elliptic curves*. Vol. 106. Springer Science & Business Media, 2009.

[27] *The Stacks project*. `https://stacks.math.columbia.edu`. 2021.

[28] Hong Bing Yu. "Note on Heath-Brown's estimate for Heilbronn's exponential sum". In: *Proc. Amer. Math. Soc.* 127.7 (1999), pp. 1995–1998.