Exponential sums over finite fields: elementary methods

E. Kowalski

ETH Zürich
Version of October 16, 2018
kowalski@math.ethz.ch

Contents

Introduction and motivation

Introduction

Although we will introduce more general examples later, we first define exponential sums over finite fields in this section as any sum of the type

$$S = \sum_{\substack{x \in \mathbf{F}_p \\ g(x) \neq 0}} \exp\Bigl(\frac{2i\pi}{p} f(x)/g(x)\Bigr)$$

where p is a prime number, $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$, and f, $g \in \mathbf{F}_p[X]$ are polynomials, such that $g \neq 0$. Here, $f(x)/g(x)$ is computed in the finite field $\mathbf{F}_p$ (and makes sense as such because x is restricted to be such that $g(x) \neq 0$). So, for instance, if $p = 7$ and $g(x) = 3 \pmod 7$, we have $1/g(x) = 5 \pmod 7$ since $3 \cdot 5 = 15 \equiv 1 \pmod 7$.

Moreover, because

$$\exp\Bigl(\frac{2i\pi\tilde{y}}{p}\Bigr)$$

is independent of the choice of an integer $\tilde{y}$ representative of $y \in \mathbf{F}_p$, these are well-defined finite sums of complex numbers, and hence $S \in \mathbf{C}$ is a complex number.

The goals of the theory are, roughly, to understand these sums. This might mean different things:

(1) Find an "explicit" ("closed-form") expression for S;

(2) Find an *upper bound* for $|S|$, which is "non-trivial"; the meaning of the last condition is of course that this bound must be better than the obvious estimate $|S| \leqslant p$, when this is possible (which means that the values $f(x)/g(x)$, $x \in \mathbf{F}_p$, are not all constant in $\mathbf{F}_p$), and the improvement is usually required (for the purpose of applications of the theory) to be of the form

$$|S| \leqslant p\theta_S^{-1} \tag{1}$$

where the saving factor satisfies $\theta_S > 1$; indeed, one often requires that θ_S be of a maximal size, as we will see later;

(3) Find a lower bound for $|S|$, if $S \neq 0$; this question is not as important in applications than the previous one, and this is fortunate, since it is in fact much harder;

(4) When S depends on further parameters t (in an arbitrary set T), try to understand the variation of S as a function of $t \in T$.

We will see examples of all these soon, but a first remark is that it is because (1) is most often an impossible target (one should compare this with the well-known fact that the indefinite integral of certain simple elementary functions – e.g., $\exp(-x^2)$ – are not themselves expressible in terms of simple operations and elementary functions) that (2), (3) and (4) naturally arise. In applications to analytic number theory (which are those we will mostly consider), Problem (2) is usually the most pressing: proving (1) for certain exponential sums, with fairly specific saving factors, is often enough to prove a highly desirable theorem.

Before going to describe "real" examples, here is the simplest exponential sum; although it is essentially trivial, its importance should not be underestimated, as the computation involved is often implicitly present in other arguments.

Example 1 (Free summation). Consider the sums

$$S_a = \sum_{x \in \mathbf{F}_p} e\Big(\frac{ax}{p}\Big)$$

where $a \in \mathbf{Z}$, p is prime, and from now on we write

$$e(z) = \exp(2i\pi z), \qquad \text{for } z \in \mathbf{C}.$$

We can compute these sums explicitly: we have

$$S_a = \begin{cases} p & \text{if } a \equiv 0 \,(\mathrm{mod}\, p) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Indeed, picking specific representatives of $\mathbf{F}_p$ in $\mathbf{Z}$, we have

$$S_a = 1 + w + w^2 + \cdots + w^{p-1}, \qquad \text{where } w = e(a/p)$$

and then we can apply the formula for a finite geometric sum, together with the fact that $w = 1$ if and only if $a/p \in \mathbf{Z}$, which means $a \equiv 0 \,(\mathrm{mod}\, p)$.

Motivation

We present, briefly and without complete details, two examples of applications of exponential sums over finite fields. The sums which occur are very important and will be considered (after being generalized) many times in this book.

Example 2 (Quadratic Gauss sums). This example is both one of the oldest to have been considered, and one of the few interesting ones where Problem (1) is solved: there is an explicit formula.

Consider the *quadratic Gauss sum*

$$G_n = \sum_{x \in \mathbf{Z}/n\mathbf{Z}} e\Big(\frac{x^2}{n}\Big),$$

where we allow any integer $n \geqslant 1$. We then have

Theorem 3 (Gauss). *For all odd integers $n \geqslant 3$, we have*

$$G_n = \begin{cases} \sqrt{n} & \text{if } n \equiv 1 \,(\mathrm{mod}\, 4) \\ i\sqrt{n} & \text{if } n \equiv 3 \,(\mathrm{mod}\, 4). \end{cases} \tag{3}$$

This result may look innocuous, but it is by no means trivial. Before giving references for its proof (we will also give the proof that $|G_n| = \sqrt{n}$ below, but not compute the actual argument), here is one of the original applications: one of the proofs of the Quadratic Reciprocity Law (this illustrates the subtlety that must be involved).

Recall first:

Definition 4. Let p be a prime number. The *Legendre symbol* modulo p is defined to be the map $\mathbf{F}_p \to \{-1,0,1\}$ denoted $x \mapsto \left(\frac{x}{p}\right)$ which is defined by

$$\left(\frac{x}{p}\right) = \begin{cases} 0 & \text{if } x = 0 \\ 1 & \text{if there exists } y \in \mathbf{F}_p \text{ such that } y^2 = x \\ -1 & \text{if there does not exist } y \in \mathbf{F}_p \text{ such that } y^2 = x. \end{cases}$$

If $n \in \mathbf{Z}$, of course, we write $\left(\frac{n}{p}\right)$ for the Legendre symbol of the reduction of n modulo p.

We then have the following result:

Theorem 5 (Gauss; the Law of Quadratic Reciprocity). *For $p \neq q$ be two odd prime numbers. We then have*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}.$$

Why is this so remarkable, or surprising? Note that if we think of $\left(\frac{p}{q}\right)$, as a function of p, it is periodic of period q, and depends only on the class of p modulo q; in particular, it seems to completely ignore the fact that p comes or not from a prime (all the more so that, as is well-known,[1] *any* non-zero class $x \in \mathbf{F}_q$ can be represented as the reduction of a prime representative $p \equiv x \pmod{q}$). On the other hand, the map $p \mapsto \left(\frac{q}{p}\right)$ seems to be of a completely different nature: its value depends on the question whether the fixed integer q is, or not, a square modulo p, with p varying. At first sight (and even at second, or third), there is *no reason* for this map to be periodic. And yet, as a consequence of Quadratic Reciprocity, it is indeed periodic (with period dividing $4q$).

Proof of Theorem 5 from Theorem 3. We need simply observe the following formula:

$$G_{pq} = G_p G_q \left(\frac{p}{q}\right)\left(\frac{q}{p}\right), \tag{4}$$

for p, q distinct odd primes, since it follows that

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \frac{G_{pq}}{G_p G_q} = \begin{cases} 1 & \text{if } p \equiv 1 \pmod 4 \text{ or } q \equiv 1 \pmod 4 \\ -1 & \text{if } p \equiv q \equiv 3 \pmod 4, \end{cases}$$

by applying Theorem 3, which is equivalent with the desired conclusion. And the proof of (4) is quite easy: by the Chinese Remainder Theorem, and the fact that $p \neq q$, so that p is invertible modulo q and conversely, we can write any $x \in \mathbf{Z}/pq\mathbf{Z}$ in a unique way as $x = px_1 + qx_2$, where x_1 is well-defined modulo q and x_2 modulo p. Thus we get

$$\begin{aligned} G_{pq} &= \sum_{x \in \mathbf{Z}/pq\mathbf{Z}} e\left(\frac{x^2}{pq}\right) \\ &= \sum_{x_1 \in \mathbf{F}_q} \sum_{x_2 \in \mathbf{F}_p} e\left(\frac{(px_1 + qx_2)^2}{pq}\right) \\ &= \Big(\sum_{x_1 \in \mathbf{F}_q} e\left(\frac{px_1^2}{q}\right)\Big) \times \Big(\sum_{x_2 \in \mathbf{F}_p} e\left(\frac{qx_2^2}{p}\right)\Big). \end{aligned} \tag{5}$$

[1] This is the famous theorem of Dirichlet on primes in arithmetic progressions.

But now we observe first that, if p is a square modulo q, say $p = y^2$, necessarily with $y \neq 0$ since $p \neq q$, we have

$$\sum_{x_1 \in \mathbf{F}_q} e\Big(\frac{px_1^2}{q}\Big) = \sum_{x_1 \in \mathbf{F}_q} e\Big(\frac{(yx_1)^2}{q}\Big) = G_q = \Big(\frac{p}{q}\Big) G_q$$

(by the change of variable $yx_1 \mapsto x$). On the other hand, if p is *not* a square modulo q, the elements px_1^2 run over all non-squares in $\mathbf{F}_q$, each of them being represented twice (since $(-x)^2 = x^2$), except for $x_1 = 0$ which represents (once) the element 0. So, using (2), we have

$$\begin{aligned}\sum_{x_1 \in \mathbf{F}_q} e\Big(\frac{px_1^2}{q}\Big) &= 1 + 2 \sum_{\substack{y \in \mathbf{F}_q^\times \\ y \text{ not a square}}} e\Big(\frac{y}{q}\Big) \\ &= 1 + 2\Big(\sum_{y \in \mathbf{F}_q^\times} e\Big(\frac{y}{q}\Big) - \sum_{\substack{y \in \mathbf{F}_q^\times \\ y \text{ a square}}} e\Big(\frac{y}{q}\Big)\Big) \\ &= 1 + 2\Big(-1 - \frac{1}{2} \sum_{x \in \mathbf{F}_q^\times} e\Big(\frac{x^2}{q}\Big)\Big) \\ &= -G_q = \Big(\frac{p}{q}\Big) G_q.\end{aligned}$$

Hence we see that, in all cases, we have

$$\sum_{x_1 \in \mathbf{F}_q} e\Big(\frac{px_1^2}{q}\Big) = \Big(\frac{p}{q}\Big) G_q$$

and applying this, and the analogue with p and q reversed, to (5), we obtain (4). □

We do not prove Theorem 3 here, since the ideas involved are largely unrelated to our purposes in this book; see, for instance, [10,] or [12,].

A last remark about this example: sums over finite rings $\mathbf{Z}/n\mathbf{Z}$, where n is not necessarily prime, are of course important in many applications; however, because the use of the Chinese Remainder Theorem mostly reduces their study to the case of n being a prime power, we will not consider them in this text, except incidentally.

Example 6 (Kloosterman sums). Kloosterman sums were first written down by Poincaré around 1912, but their first application arose when Kloosterman introduced them independently in the 1920's (see [15] for a survey of the story of Kloosterman sums in analytic number theory). Their definition – we allow arbitrary modulus here – is as follows:

Definition 7 (Kloosterman sum). Let $c \geqslant 1$ be an integer, $m, n \in \mathbf{Z}$. The associated Kloosterman sum $S(m, n; c)$ is defined by

$$S(m, n; c) = \sum_{\substack{x \in \mathbf{Z}/c\mathbf{Z} \\ (x,c)=1}} e\Big(\frac{mx + n/x}{c}\Big)$$

were $1/x$ is the inverse in $\mathbf{Z}/c\mathbf{Z}$ of the invertible element $x \in (\mathbf{Z}/c\mathbf{Z})^\times$.

The definition is probably un-enlightening when first encountered. But here is the first application that Kloosterman derived from studying these sums, a beautiful result which is hopefully of clear arithmetic significance.

THEOREM 8 (Kloosterman). *Let $a_1, \ldots, a_4 \geqslant 1$ be positive integers, and let $n \geqslant 1$ be a positive integer. Then, for all n large enough, depending on the a_i's, there exists at least one integral solution $(x_1, \ldots, x_4) \in \mathbf{Z}^4$ to the diophantine equation*

$$a_1x_1^2 + \cdots + a_4x_4^2 = n, \tag{6}$$

provided there is no congruence obstruction.

In fact, Kloosterman's statement is much more precise (it gives for instance an asymptotic formula for the number of integral solutions of (6); see, e.g., [12,] for a modern treatment).

Although seemingly unrelated, the most crucial ingredient (not the only one) in the proof of this theorem was a non-trivial estimate for Kloosterman sums. Precisely, Kloosterman proved

THEOREM 9. *Let p be a prime number and n, m integers coprime with p. We then have the upper bound*

$$|S(n,m;p)| \leqslant 2p^{3/4}.$$

This is a first instance of Problem (2) (bounding exponential sums), and it is quite successful: the saving factor here is of size $\theta_S \approx p^{1/4}$ for $S = S(m,n;p)$.

The proof of this theorem will be given in the next chapter (see Section 2.2). The argument is quite nice and contains ideas that are still of use. However, the result itself can be improved, and one of the main results of this book, the Riemann Hypothesis for one-variable sums over finite fields (due to A. Weil), will imply the following improvement:

THEOREM 10 (Weil). *Let p be a prime number and n, m integers coprime with p. We then have the upper bound*

$$|S(n,m;p)| \leqslant 2\sqrt{p}.$$

We will also see (as a consequence of the work required for the proof of Theorem 9) that this is in some sense best possible: for any prime p, Kloosterman showed that there exists m, $n \in \mathbf{F}_p^\times$ such that

$$|S(m,n;p)| > \sqrt{2p-2} \tag{7}$$

(see again Section 2.2 for a proof).

However, one should not imagine that Kloosterman sums are so well understood. There are easy-looking questions which remain quite out of reach, e.g.:

QUESTION. Are there infinitely many prime numbers p such that $S(1,1;p) > 0$?

Notation

Most notation is very standard, and we only summarize here the most common. We write $|X|$ for the cardinality of a set, and in particular $|X| = +\infty$ means that X is infinite, with no indication on the infinite cardinal involved.

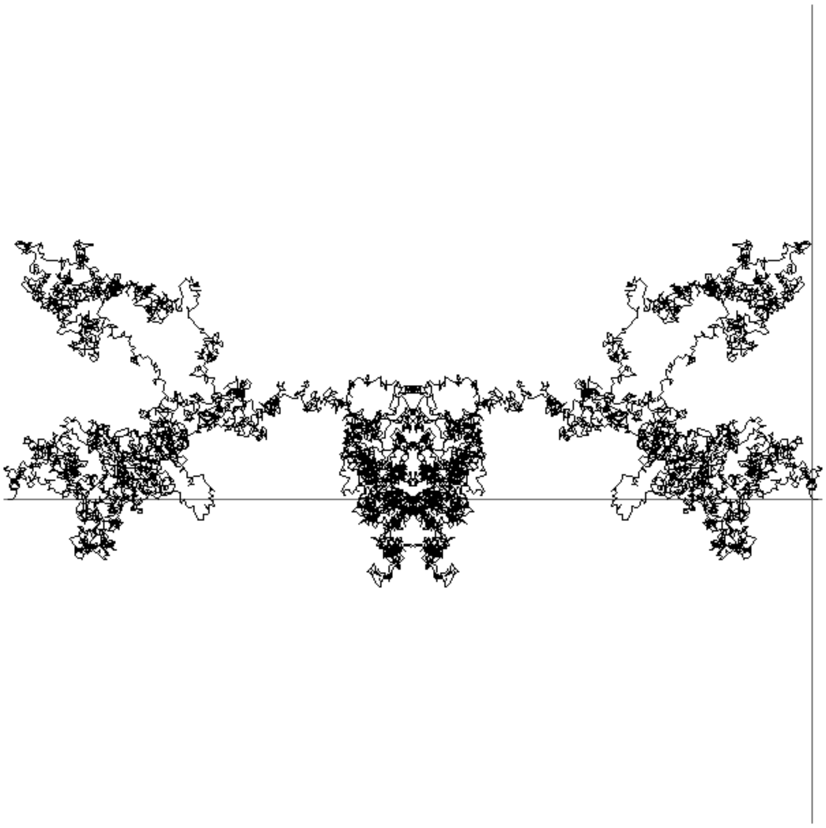
As already indicated, we write $e(z) = \exp(2i\pi z)$ for $z \in \mathbf{C}$; we then have $e(z+w) = e(z)e(w)$ for all z, $w \in \mathbf{C}$, $e(z+m) = e(z)$ if $m \in \mathbf{Z}$, and $e(z) = 1$ if and only if $z \in \mathbf{Z}$.

By $f \ll g$ for $x \in X$, or $f = O(g)$ for $x \in X$, where X is an arbitrary set on which f is defined, we mean synonymously that there exists a constant $C \geqslant 0$ such that $|f(x)| \leqslant Cg(x)$ for all $x \in X$. The "implied constant" is any admissible value of C. It may depend on the set X which is always specified or clear in context. The notation $f \asymp g$ means $f \ll g$ and $g \ll f$. On the other hand $f(x) = o(g(x))$ as $x \to x_0$ is a topological statement meaning that $f(x)/g(x) \to 0$ as $x \to x_0$.

We conclude this introduction with a graphical illustration of the complexity of exponential sums: Figure 1 shows the path in the complex plane formed by starting from the origin and connecting with line segments the successive partial sums

$$\sum_{1\leqslant x\leqslant n} e\Bigl(\frac{x+1/x}{p}\Bigr)$$

for $1 \leqslant n \leqslant p-1$, for the specific value $p = 10007$.



The path of the partial sums of $S(1,1;10007) = -151.358543\ldots$

Acknowledgments. Thanks to T. Browning, K. Gong, M. Einsiedler, M. Balazard, O. Marfaing, B. Tackmann and others for various corrections to the text.

CHAPTER 1

Finite fields and characters

This chapter is mostly concerned with introducing material about finite fields and characters of finite abelian groups, which many readers have probably already encountered – they may skip to the next chapter without loss in that case.

1.1. Reminders on finite fields

Here are, quite quickly sketched, the fundamental facts about finite fields; it is expected that there will not be much new here for most readers, but we refer to [10,] for a complete account.

(1) For $q \geqslant 1$, there exists a field with order q if and only if $q = p^\nu$ is a power of a prime p with $\nu \geqslant 1$ (note that rumors about the existence of a field with 1 element are much exaggerated). This finite field is unique up to isomorphism, but *not* up to unique isomorphism (see Remark 1.1 below). Taking due care to not claim that there is *a* field with q elements, we will usually denote by $\mathbf{F}_q$ a chosen field with q elements. Unless specified otherwise, the associated prime number is denoted p; it can be recovered as the characteristic of $\mathbf{F}_q$, i.e., we have

$$p\mathbf{Z} = \{n \in \mathbf{Z} \mid nx = 0 \text{ for all } x \in \mathbf{F}_q\}.$$

For $\nu = 1$, we have a *canonical* isomorphism $\mathbf{F}_p \simeq \mathbf{Z}/p\mathbf{Z}$ (by sending the unit element 1 to 1); in general, such a field $\mathbf{F}_q$ is an $\mathbf{F}_p$-vector space of dimension $\nu = [\mathbf{F}_q : \mathbf{F}_p]$.

(2) Given a prime number p, all finite fields of characteristic p (i.e., of p-power order) can be recovered up to isomorphism by fixing an algebraic closure $\bar{\mathbf{F}}_p$ of the *prime field* $\mathbf{F}_p$; then, for each $\nu \geqslant 1$, $\bar{\mathbf{F}}_p$ contains a unique subfield of order $q = p^\nu$, given by

$$\mathbf{F}_q = \{x \in \bar{\mathbf{F}}_p \mid x^q = x\} \subset \bar{\mathbf{F}}_p \ ;$$

this is also the splitting field in $\bar{\mathbf{F}}_p$ of the polynomial $X^q - X \in \mathbf{F}_p[X]$.

For each such subfield $\mathbf{F}_q$ and any $\nu \geqslant 1$, there is a unique extension of $\mathbf{F}_q$ of degree ν contained in $\bar{\mathbf{F}}_p$, namely the field $\mathbf{F}_{q^\nu}$. Moreover, $\bar{\mathbf{F}}_p$ is also an algebraic closure of $\mathbf{F}_q$.

We have then inclusions between these fields determined by $\mathbf{F}_{q^\nu} \subset \mathbf{F}_{q^\mu}$ if and only if $\nu \mid \mu$.

(3) For any finite field $\mathbf{F}_q$ with q elements and extension field (finite or infinite) $k/\mathbf{F}_q$, the map

$$\mathrm{Fr}_q : \begin{cases} k & \longrightarrow & k \\ x & \mapsto & x^q \end{cases}$$

is a field automorphism of k such that $\mathbf{F}_q$ is the fixed field of Fr_q; it is called the *arithmetic Frobenius* of $\mathbf{F}_q$.

(4) Any finite field extension $\mathbf{F}_{q^\nu}/\mathbf{F}_q$ of degree ν is a Galois extension, and its Galois group is cyclic of order ν generated by the Frobenius; in other words, there is a

canonical isomorphism

$$\begin{cases} \mathbf{Z}/\nu\mathbf{Z} & \xrightarrow{\sim} & \mathrm{Gal}(\mathbf{F}_{q^\nu}/\mathbf{F}_q) \\ 1 & \mapsto & \mathrm{Fr}_q \end{cases};$$

in particular, by Galois theory, we recover the formula

$$\mathbf{F}_q = \{x \in \mathbf{F}_{q^\nu} \mid x^q = x\}, \tag{1.1}$$

already noted earlier. Note that, as is customary, we will write either x^σ or $\sigma(x)$ for the image of an element x in any field under an automorphism σ.

(5) (For readers familiar with infinite Galois theory; this will not be of much use in this book) For any finite field $\mathbf{F}_q$ with q elements and choice of an algebraic closure $\bar{\mathbf{F}}_q$, the latter is also a separable closure of $\mathbf{F}_q$, and there is a canonical isomorphism

$$\begin{cases} \hat{\mathbf{Z}} & \xrightarrow{\sim} & \mathrm{Gal}(\bar{\mathbf{F}}_q/\mathbf{F}_q) \\ 1 & \mapsto & \mathrm{Fr}_q \end{cases}$$

Among these facts, the formula (1.1) is the key fact that allows algebraic methods to be useful in the theory of exponential sums over finite fields. The reader will, for instance, see concretely how it comes into play as a *deus ex machina* in the Stepanov method.

Remark 1.1. We have remarked above that a finite field with q elements is not defined canonically if q is not a prime number (in which case, the identification with $\mathbf{Z}/q\mathbf{Z}$ is canonical). This is due to the presence of automorphisms of $\mathbf{F}_q$ – namely the Frobenius and its powers. Concretely, this means the following for instance: suppose we have two number fields $K_1/\mathbf{Q}$, $K_2/\mathbf{Q}$ and a prime number p, unramified for both of them, and prime ideals $\mathfrak{p}_1$, $\mathfrak{p}_2$ above p in K_1 and K_2 respectively. If $N\mathfrak{p}_1 = N\mathfrak{p}_2$, we know that the residue fields k_1, k_2 at $\mathfrak{p}_1$, $\mathfrak{p}_2$ have the same order, hence are isomorphic. *It is not obvious at all* how to construct an explicit isomorphism between k_1 and k_2 (in terms, say, of polynomials defining K_1 and K_2).

Besides these foundational facts, we need to know the definition and properties of the norm and trace maps that related two finite fields, one of which is an extension of the other.

Definition 1.2 (Norm and trace). Let $\mathbf{F}_{q^\nu}/\mathbf{F}_q$ be a finite extension of finite fields with the indicated orders. The *trace map* from $\mathbf{F}_{q^\nu}$ to $\mathbf{F}_q$ is the $\mathbf{F}_q$-linear map

$$\mathrm{Tr} = \mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q} : \begin{cases} \mathbf{F}_{q^\nu} & \longrightarrow & \mathbf{F}_q \\ x & \mapsto & x + x^q + \cdots + x^{q^{\nu-1}} \end{cases}$$

and the *norm map* is the multiplicative map

$$N = N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q} : \begin{cases} \mathbf{F}_{q^\nu} & \longrightarrow & \mathbf{F}_q \\ x & \mapsto & x \cdot x^q \cdots x^{q^{\nu-1}} = x^{(q^\nu-1)/(q-1)} \end{cases}$$

which restricts to a multiplicative group-homomorphism

$$N : \mathbf{F}_{q^\nu}^\times \longrightarrow \mathbf{F}_q^\times.$$

We will usually avoid spelling out which fields are involved when this is clear from context.

If we write

$$\mathrm{Tr}(x) = \sum_{\sigma \in \mathrm{Gal}(\mathbf{F}_{q^\nu}/\mathbf{F}_q)} x^\sigma, \quad N(x) = \prod_{\sigma \in \mathrm{Gal}(\mathbf{F}_{q^\nu}/\mathbf{F}_q)} x^\sigma,$$

we see clearly that

$$\sigma(\mathrm{Tr}(x)) = \mathrm{Tr}(x), \qquad \sigma(N(x)) = N(x)$$

for all elements σ of the Galois group, and the implied fact that the trace and norm take values in $\mathbf{F}_q$ follows. We also find similarly that

$$\mathrm{Tr}(\sigma(x)) = \mathrm{Tr}(x), \quad N(\sigma(x)) = N(x), \tag{1.2}$$

in particular

$$\mathrm{Tr}(x^q) = \mathrm{Tr}(x), \quad N(x^q) = N(x).$$

These, it turns out, are the only ways to obtain the same trace or norm.

LEMMA 1.3. *Let $\mathbf{F}_{q^\nu}/\mathbf{F}_q$ be a finite extension of finite fields with the indicated orders.*

(1) *The trace map is a surjective linear map*

$$\mathbf{F}_{q^\nu} \xrightarrow{\mathrm{Tr}} \mathbf{F}_q,$$

with

$$\mathrm{Ker}(\mathrm{Tr}) = \{x \in \mathbf{F}_{q^\nu} \mid x = y^q - y \textit{ for some } y \in \mathbf{F}_{q^\nu}\}.$$

In particular, there exists $x_0 \in \mathbf{F}_{q^\nu}$ such that $\mathrm{Tr}(x_0) \neq 0$.

(2) *The norm map is a surjective homomorphism*

$$\mathbf{F}_{q^\nu}^\times \xrightarrow{N} \mathbf{F}_q^\times,$$

with

$$\mathrm{Ker}(N) = \{x \in \mathbf{F}_{q^\nu}^\times \mid x = y^{q-1} \textit{ for some } y \in \mathbf{F}_{q^\nu}^\times\}. \tag{1.3}$$

There exists $x_0 \in \mathbf{F}_{q^\nu}^\times$ such that $N(x_0) \neq 1$ if and only if $q \neq 2$.

PROOF. Although these are facts which hold in greater generality (e.g., there are analogues for all finite cyclic Galois extensions), we give a short proof illustrating how special features of finite fields can be very useful.

For (1), define

$$\delta : \begin{cases} \mathbf{F}_{q^\nu} & \longrightarrow & \mathbf{F}_{q^\nu} \\ y & \mapsto & y^q - y \end{cases},$$

which is an $\mathbf{F}_q$-linear map with

$$\mathrm{Ker}(\delta) = \{y \mid y^q = y\} = \mathbf{F}_q,$$

and hence – in particular – we have $\dim \mathrm{Im}(\delta) = \nu - 1$. We want to show that $\mathrm{Ker}(\mathrm{Tr}) = \mathrm{Im}(\delta)$, and this will be enough to obtain all the results because it follows then that $\dim \mathrm{Ker}(\mathrm{Tr}) = \nu - 1$, hence $\dim \mathrm{Im}(\mathrm{Tr}) = 1$, showing the surjectivity of the trace.

We have already observed that $\mathrm{Ker}\,\mathrm{Tr} \supset \mathrm{Im}(\delta)$; but on the other hand, we can write[1]

$$\mathrm{Ker}(\mathrm{Tr}) = \{x \in \mathbf{F}_{q^\nu} \mid P(x) = 0\}$$

where P is a polynomial of degree $q^{\nu-1}$ given by

$$P = X + X^q + \cdots + X^{q^{\nu-1}}$$

hence, from field theory, we know that

$$|\mathrm{Ker}(\mathrm{Tr})| \leqslant \deg(P) \leqslant q^{\nu-1},$$

and by the inclusion already known, there must be equality.

[1] This is where the nature of finite fields plays a role; usually the trace is not a uniform polynomial of the argument.

The argument for (2) is quite similar; the analogue of δ is the group homomorphism

$$\Delta : \begin{cases} \mathbf{F}_{q^\nu}^\times & \longrightarrow & \mathbf{F}_{q^\nu}^\times \\ y & \mapsto & y^{q-1} \end{cases},$$

with $\operatorname{Ker}(\Delta) = \mathbf{F}_q^\times$, hence $|\operatorname{Im}(\Delta)| = (q^\nu - 1)/(q-1)$. We have also $\operatorname{Im}(\Delta) \subset \operatorname{Ker}(N)$, and we must show equality, and this follows from the remark that

$$\operatorname{Ker}(N) = \{x \in \mathbf{F}_{q^\nu}^\times \mid Q(x) = 0\}, \qquad Q = X^{(q^\nu-1)/(q-1)} - 1,$$

so that $|\operatorname{Ker}(N)| \leqslant (q^\nu - 1)/(q-1)$. Therefore, we have $|\operatorname{Im}(N)| = (q^\nu - 1)/|\operatorname{Ker}(N)| = q - 1 = |\mathbf{F}_q^\times|$, proving the surjectivity of the norm.

Finally the last remark is clear, since $\mathbf{F}_q^\times = \{1\}$ if and only if $q = 2$. □

Remark 1.4. In other words, we have shown that we have the following two exact sequences of abelian groups:

$$0 \to \mathbf{F}_q \hookrightarrow \mathbf{F}_{q^\nu} \xrightarrow{\delta} \mathbf{F}_{q^\nu} \xrightarrow{\operatorname{Tr}} \mathbf{F}_q \to 0,$$

and

$$1 \to \mathbf{F}_q^\times \hookrightarrow \mathbf{F}_{q^\nu}^\times \xrightarrow{\Delta} \mathbf{F}_{q^\nu}^\times \xrightarrow{N} \mathbf{F}_q^\times \to 1.$$

A common way to "write down" an element x in an extension field $\mathbf{F}_{q^\nu}$ of $\mathbf{F}_q$ is to specify a polynomial $0 \neq f \in \mathbf{F}_q[X]$ of which it is a root. It is quite useful that one can write down the trace and norm of x in terms of f.

Lemma 1.5. *Let $\mathbf{F}_q$ be a finite field, let $f \in \mathbf{F}_q[X]$ be a non-zero irreducible monic polynomial of degree $d \geqslant 1$. Then, for any $x \in \mathbf{F}_{q^d}$ such that $f(x) = 0$, we have*

$$\operatorname{Tr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(x) = -a_{d-1}, \qquad N_{\mathbf{F}_{q^d}/\mathbf{F}_q}(x) = (-1)^d a_0, \tag{1.4}$$

where

$$f = X^d + a_{d-1}X^{d-1} + \cdots + a_1 X + a_0, \qquad a_i \in \mathbf{F}_q.$$

Proof. Since f is irreducible, the set $\{x^{q^j}\}$, $0 \leqslant j \leqslant d-1$, of Galois conjugates of x is identical with the set of zeros of f. Since f is monic, it follows that we can factor

$$f = \prod_{\sigma \in \operatorname{Gal}(\mathbf{F}_{q^d}/\mathbf{F}_q)} (X - \sigma(x)),$$

and comparing with the expansion in powers of X, we see that

$$\operatorname{Tr}(x) = \sum_{\sigma \in \operatorname{Gal}(\mathbf{F}_{q^d}/\mathbf{F}_q)} \sigma(x)$$

is the negative of the coefficient of X^{d-1}, while

$$N(x) = \prod_{\sigma \in \operatorname{Gal}(\mathbf{F}_{q^d}/\mathbf{F}_q)} \sigma(x)$$

is $(-1)^d$ times the value at 0, i.e., times the constant coefficient. □

The last fact we need is another result essentially due to Gauss:

Lemma 1.6. *Let $\mathbf{F}_q$ be a finite field with q elements. Then the multiplicative group $\mathbf{F}_q^\times$ is cyclic of order $q-1$. In particular, there are*[2] *$\varphi(q-1)$ generators of $\mathbf{F}_q^\times$; these are called* primitive roots *in $\mathbf{F}_q$.*

[2] Here, $\varphi(n)$ denotes the Euler function, i.e., the number of invertible elements in $\mathbf{Z}/n\mathbf{Z}$.

Proof. We use the structure theory of finite abelian groups to get a quick result: there exist integers $k \geqslant 1$ and $d_1, \ldots, d_k \geqslant 1$ such that

$$\mathbf{F}_q^\times \simeq \mathbf{Z}/d_1\mathbf{Z} \times \mathbf{Z}/d_1 d_2\mathbf{Z} \times \cdots \times \mathbf{Z}/d_1 \cdots d_k\mathbf{Z}$$

and we observe that, in the group on the right-hand side, the equation (multiplicatively written)

$$x^{d_1} = 1$$

has d_1^k solutions, namely the elements of

$$\mathbf{Z}/d_1\mathbf{Z} \times d_2\mathbf{Z}/d_1 d_2\mathbf{Z} \times \cdots \times (d_2 \cdots d_k)\mathbf{Z}/(d_1 \cdots d_k)\mathbf{Z}.$$

On the other hand, this equation is a polynomial equation $P(x) = 0$ in $\mathbf{F}_q$, with $P = X^{d_1} - 1$, and therefore it has $\leqslant d_1$ solutions. It follows that $d_1^k \leqslant d_1$, which implies $k = 1$, and $\mathbf{F}_q^\times \simeq \mathbf{Z}/d_1\mathbf{Z}$ is cyclic. □

Remark 1.7. Using the formula

$$\varphi(n) = n \prod_{\ell \mid n} (1 - \ell^{-1}),$$

(where ℓ runs over primes dividing n) and the fact that

$$\lim_{x \to +\infty} \prod_{\ell \leqslant x} (1 - \ell^{-1}) = 0,$$

it is not difficult to check that

$$\liminf_{q \to +\infty} \frac{\varphi(q-1)}{q-1} = \liminf_{q \to +\infty} \prod_{\ell \mid q-1} (1 - \ell^{-1}) = 0,$$

so that the proportion of primitive roots among elements of $\mathbf{F}_q^\times$ may be arbitrarily small.

The problem of finding, efficiently, a primitive root in $\mathbf{F}_q^\times$ is a difficult one: it is not known how to do it in polynomial time (with respect to the number of digits of q, or equivalently with respect to $\log q$).

1.2. Characters of finite abelian groups

The content of this section is, again, likely to be well-known. As in the previous section, we proceed quickly to prove the main facts, without trying to be economical in the proofs.

We have used the map $x \mapsto e(x/p)$ on $\mathbf{F}_p$ to define exponential sums; its main feature, that leads to generalizations over other finite fields, is that this is a group homomorphism with complex values. Such homomorphisms are called *characters*.

Definition 1.8 (Character of a group). Let G be any group. A *character* of G is a group homomorphism

$$\chi : G \longrightarrow \mathbf{C}^\times.$$

The following facts are quite obvious:

(1) The set $\hat{G}$ of characters of G is a group with the pointwise multiplication

$$(\chi_1 \cdot \chi_2)(x) = \chi_1(x)\chi_2(x),$$

the inverse defined by $(\chi^{-1})(x) = \chi(x)^{-1}$, and the unit the trivial homomorphism $x \mapsto 1$.

(2) If G is a finite group of order n, all characters take values in the set

$$\boldsymbol{\mu}_n = \{z \in \mathbf{C} \mid z^n = 1\},$$

of n-th roots of unity in $\mathbf{C}$, and the inverse χ^{-1} of a character is also its complex conjugate $\bar{\chi}$.

Example 1.9. (1) The map $z \mapsto e(z)$ is a character of the additive group of $\mathbf{C}$.

(2) For any fixed $n \in \mathbf{Z}$, the map $x \mapsto e(nx)$ is a character of the quotient group $\mathbf{R}/\mathbf{Z}$; in fact (under minimal regularity assumptions, such as looking at measurable characters only) we have $\widehat{\mathbf{R}/\mathbf{Z}} \simeq \mathbf{Z}$.

The main result concerning characters of finite abelian group is the following:

Proposition 1.10 (Characters of finite abelian groups). *Let G be a finite abelian group. Then the set of characters of G forms an orthonormal basis of the space*

$$C(G) = \{f\,:\,G \longrightarrow \mathbf{C}\}$$

of complex-valued functions on G with respect to the inner-product

$$\langle f, g\rangle = \frac{1}{|G|}\sum_{x\in G} f(x)\overline{g(x)}.$$

In particular, one deduces immediately the following facts from this proposition; they are used constantly (indeed, some have already been implicitly used in the introductory chapter!).

(1) The dual group $\hat{G}$ is of order n; in fact, one can show (or indeed see from the proof of the proposition) that $\hat{G}$ is isomorphic to G, but such isomorphisms are not canonical and not particularly useful in general.

(2) We have the *orthogonality relations*:

$$\sum_{x\in G}\chi_1(x)\overline{\chi_2(x)} = \begin{cases} |G| & \text{if } \chi_1 = \chi_2 \\ 0 & \text{otherwise,}\end{cases} \tag{1.5}$$

for all χ_1, $\chi_2 \in \hat{G}$, and

$$\sum_{\chi\in\hat{G}}\chi(x)\overline{\chi(y)} = \begin{cases} |G| & \text{if } x = y \\ 0 & \text{otherwise.}\end{cases} \tag{1.6}$$

Indeed, the first one is the direct translation of the orthonormality of the characters, and the second can be seen either as stating that the transpose of a unitary matrix is unitary, or as the expansion in the basis of characters of δ functions: fixing y in G, the function

$$\delta_y\,:\,x \mapsto \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise,}\end{cases}$$

has the expansion

$$\delta_y(x) = \sum_{\chi\in\hat{G}}\langle\delta_y, \chi\rangle\chi(x)$$

and the coefficients are given by

$$\langle\delta_y, \chi\rangle = \frac{1}{|G|}\sum_{x\in G}\delta_y(x)\overline{\chi(x)} = \frac{1}{|G|}\overline{\chi(y)},$$

so that the expansion corresponds exactly to (1.6). As we will see on numerous occasions, this second orthogonality formula is very useful to "transcribe" analytically a constraint $x = y$ appearing in a problem.

Proof of Proposition 1.10. To give a quick proof, we use the structure theory of finite abelian groups; this implies that we have an isomorphism

$$G \simeq \mathbf{Z}/d_1\mathbf{Z} \times \cdots \times \mathbf{Z}/d_k\mathbf{Z},$$

where $k \geqslant 1$ and the integers $d_1, \ldots, d_k$ are uniquely determined by G. However, for abelian groups G_1 and G_2, there is a canonical isomorphism

$$\begin{cases} \hat{G}_1 \times \hat{G}_2 & \xrightarrow{\sim} & \widehat{G_1 \times G_2} \\ (\chi_1, \chi_2) & \mapsto & (x, y) \mapsto \chi_1(x)\chi_2(y) \end{cases}$$

(the only non-obvious thing is the surjectivity, but this follows from

$$\chi(x, y) = \chi((x, 1) \cdot (1, y)) = \chi((x, 1)) \cdot \chi((1, y))$$

for any $\chi \in \widehat{G_1 \times G_2}$, where the two factors clearly define characters of G_1 and G_2, respectively, forming a pair which maps to χ under the map above). This map, and the analogue for multiple factors, are compatible with the Hilbert space structure in the following sense: there is an isomorphism

$$\begin{cases} C(G_1) \otimes C(G_2) & \xrightarrow{\sim} & C(G_1 \times G_2) \\ f \otimes g & \mapsto & (x, y) \mapsto f(x)g(y) \end{cases}$$

such that

$$\langle f \otimes g, f_1 \otimes g_1 \rangle = \langle f, f_1 \rangle \langle g, g_1 \rangle,$$

for $f_i \in C(G_i)$, $g_i \in C(G_i)$. This implies that if the characters of G_1 and G_2 form orthonormal bases of $C(G_1)$ and $C(G_2)$, the same property holds for those of $G_1 \times G_2$ and $C(G_1 \times G_2)$. In other words, using the structure theorem, we need only check the result when G is cyclic, say $G = \mathbf{Z}/d\mathbf{Z}$ with $d \geqslant 1$.

In that case, since the group can be described as the group generated by a single element a satisfying the only relation $a^d = 1$, it follows that mapping χ to $\chi(1) \in \mathbf{C}^\times$ gives an isomorphism

$$\hat{G} = \mathrm{Hom}(G, \mathbf{C}^\times) \simeq \{z \in \mathbf{C}^\times \mid z^d = 1\} = \boldsymbol{\mu}_d \simeq \mathbf{Z}/d\mathbf{Z},$$

so that the characters of G are the maps $x \mapsto \zeta^x$ where ζ runs over $\boldsymbol{\mu}_d$. We can write $\zeta = e(a/d)$, for a unique $a \in \mathbf{Z}/d\mathbf{Z}$, and then the characters take the already familiar form

$$e_a : x \mapsto e\Big(\frac{ax}{d}\Big), \qquad a \in \mathbf{Z}/d\mathbf{Z}.$$

Then the orthonormality becomes a simple check:

$$\begin{aligned} \langle e_a, e_b \rangle &= \frac{1}{d} \sum_{0 \leqslant x \leqslant d-1} e_a(x)\overline{e_b(x)} \\ &= \frac{1}{d} \sum_{0 \leqslant x \leqslant d-1} e\Big(\frac{(a-b)x}{d}\Big) \\ &= \delta(a, b) \end{aligned}$$

by the same geometric sum argument seen in Example 1. □

Remark 1.11. The orthonormality can also be proved directly for any group G by writing

$$\langle \chi_1, \chi_2 \rangle = \frac{1}{|G|} \sum_{x \in G} \chi(x), \qquad \chi = \chi_1 \bar{\chi_2},$$

and observing that, if $\chi \neq 1$ (i.e., $\chi_1 \neq \chi_2$), there exists an $x_0 \in G$ with $\chi(x_0) \neq 1$; then the bijective substitution $x = x_0 y$ leads to

$$\langle \chi_1, \chi_2 \rangle = \frac{1}{|G|} \sum_{y \in G} \chi(x_0 y) = \chi(x_0) \langle \chi_1, \chi_2 \rangle$$

which then gives $\langle \chi_1, \chi_2 \rangle = 0$. The case $\chi_1 = \chi_2$ is immediate.

Our interest in characters revolves around finite fields. Given a finite field $\mathbf{F}_q$ with q elements, there are actually two groups – of a quite different nature – to consider: the additive group $(\mathbf{F}_q, +)$, and the multiplicative group $(\mathbf{F}_q^\times, \cdot)$, and it is customary to speak of *additive* or *multiplicative* characters to discuss them.

Example 1.12 (Additive characters of finite fields). If $\mathbf{F}_q$ is a finite field with $q = p^\nu$ elements, we have an isomorphism of groups $\mathbf{F}_q \simeq \mathbf{F}_p^\nu \simeq (\mathbf{Z}/p\mathbf{Z})^\nu$. Since the characters of $\mathbf{Z}/p\mathbf{Z}$ are very explicitly known (as seen in the proof of Proposition 1.10), it is quite simple to give a uniform description of additive characters.

Proposition 1.13. *Let $\mathbf{F}_q$ be a finite field of characteristic p with q elements. Then there is an isomorphism*

$$\begin{cases} \mathbf{F}_q & \longrightarrow & \widehat{\mathbf{F}_q} \\ a & \mapsto & \psi_a \end{cases}$$

where the character ψ_a is defined by

$$\psi_a(x) = e\Big(\frac{\mathrm{Tr}_{\mathbf{F}_q/\mathbf{F}_p}(ax)}{p}\Big).$$

More generally, if ψ is an additive character of $\mathbf{F}_q$ and $\mathbf{F}_{q^\nu}/\mathbf{F}_q$ is a finite field extension, the map

$$x \mapsto \psi(\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(x))$$

is a character of $\mathbf{F}_{q^\nu}$.

Example 1.14 (Multiplicative characters of finite fields). Although $\mathbf{F}_q^\times$ is a cyclic group (Lemma 1.6), which might suggest a very simple structure, the group of multiplicative characters of a finite field is, in fact, quite a complicated object. The reason is that, as we already mentioned, it is quite difficult to find an *explicit* isomorphism $\mathbf{F}_q^\times \simeq \mathbf{Z}/(q-1)\mathbf{Z}$, as would be required to easily construct the characters of $\mathbf{F}_q^\times$.

Still, one may at least provide some theoretical information; in particular, note that if χ is a multiplicative character of $\mathbf{F}_q^\times$, its order is a divisor of $q-1$. Indeed, if $d \mid q-1$ is a divisor of $q-1$, the structure of cyclic groups shows that the characters of order dividing d form themselves a subgroup of order d.

In particular, if q is odd, we have $2 \mid q-1$, so that there must be a unique non-trivial character χ_2 of order 2 (the other of order dividing 2 is the trivial one); because its values lie in $\boldsymbol{\mu}_2 = \{-1, 1\}$, the only group of roots of unity to be included in $\mathbf{R}$, this is also called a (or "the") real character of $\mathbf{F}_q^\times$. For $q = p \geqslant 3$ an odd prime, we've already met it: it is the Legendre symbol (definition 4), since we have

$$\chi_2(x) = \Big(\frac{x}{p}\Big), \qquad \text{for all } x \in \mathbf{F}_p^\times = (\mathbf{Z}/p\mathbf{Z})^\times.$$

Indeed, notice that χ_2, being of order 2, must be *trivial* (equal to 1) on the subgroup $(\mathbf{F}_p^\times)^2$ of all squares x^2 with $x \in \mathbf{F}_p^\times$; hence the homomorphism χ_2 must factor

$$\mathbf{F}_p^\times \to \mathbf{F}_p^\times/(\mathbf{F}_p^\times)^2 \simeq \{-1, 1\},$$

(where the last isomorphism is due simply to the fact that the middle group is of order 2; $(\mathbf{F}_p^\times)^2$ is of order $(p-1)/2$ as the image of the homomorphism $x \mapsto x^2$ with kernel $\{\pm 1\}$), and hence χ_2 must map non-squares to -1. This is exactly the recipe to compute the Legendre symbol.

Another useful fact later on will be the analogue of the last part of Proposition 1.13:

Lemma 1.15. *Let $\mathbf{F}_{q^\nu}/\mathbf{F}_q$ be a finite extension of finite fields. For any character χ of $\mathbf{F}_q^\times$, the composite*

$$x \mapsto \chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(x))$$

is a character of $\mathbf{F}_{q^\nu}^\times$, of order equal to the order of χ.

The last part is immediate from the surjectivity of the norm map (Lemma 1.3). Note that, for the non-trivial real character χ_2 of $\mathbf{F}_q^\times$, for q odd, it follows (by its unicity) that

$$\chi_2(x) = \Bigl(\frac{N_{\mathbf{F}_q/\mathbf{F}_p}(x)}{p}\Bigr).$$

As an application of (1.6) in this context, we note that the following formula: for any d dividing $q-1$ and $x \in \mathbf{F}_q^\times$, we have

$$\sum_{\chi^d=1} \chi(x) = \begin{cases} d & \text{if } x \text{ is a } d\text{-th power, } x = y^d \text{ for some } y \in \mathbf{F}_q^\times, \\ 0 & \text{otherwise} \end{cases}, \tag{1.7}$$

(the sum running over all characters of order dividing d; this can be used to detect analytically a condition that an element is a d-th power). Indeed, one notes that the characters of the quotient group $\mathbf{F}_q^\times/(\mathbf{F}_q^\times)^d$ correspond to characters of $\mathbf{F}_q^\times$ which are of order dividing d, by the composites

$$\mathbf{F}_q^\times \to \mathbf{F}_q^\times/(\mathbf{F}_q^\times)^d \xrightarrow{\chi} \mathbf{C}^\times,$$

and since x is trivial in this quotient group if and only if it is a d-th power, the formula is indeed a particular case of (1.6).

A last remark is that, as in the case of the Legendre character, it is often very useful to define $\chi(0) = 0$ for a non-trivial multiplicative character of $\mathbf{F}_q^\times$, while $\chi(0) = 1$ if χ is the trivial character. With this convention, the last formula can be generalized to

$$\sum_{\chi^d=1} \chi(x) = |\{y \in \mathbf{F}_q \mid y^d = x\}|, \tag{1.8}$$

and it is now valid for all $d \mid q-1$ and $x \in \mathbf{F}_q$ (each side being equal to 1 for $x = 0$). Moreover, we retain the multiplicativity

$$\chi(xy) = \chi(x)\chi(y) \tag{1.9}$$

for all x, $y \in \mathbf{F}_q$ (when $x = 0$, this translates to $\chi(0) = \chi(0)\chi(x)$, which is always true).

Example 1.16 ("General" character sums). We can now define quite general exponential sums in one variable over a finite field (more general ones will only be mentioned incidentally in this book). Let $\mathbf{F}_q$ be a finite field with q elements; we assume given an additive character ψ of $\mathbf{F}_q$, and a multiplicative character χ of $\mathbf{F}_q^\times$, as well as polynomials f_1, g_1, f_2, $g_2 \in \mathbf{F}_q[X]$, with g_1, $g_2 \neq 0$. Then we denote

$$U(\mathbf{F}_q) = \{x \in \mathbf{F}_q \mid f_1(x),\ g_1(x),\ g_2(x) \neq 0\},$$

and we define

$$S = \sum_{x \in U(\mathbf{F}_q)} \chi(f_1(x)/g_1(x))\psi(f_2(x)/g_2(x)). \tag{1.10}$$

Such sums, called "general character sums", turn out to be fairly ubiquitous in analytic number theory. Their study – and that of their various generalizations – is one of the highlights of modern number theory, and in particular of its interactions with algebraic geometry...

CHAPTER 2

Elementary examples

In this chapter, we break the monotony of the previous one by giving examples (the first two of which are related to those used in the first motivation section) of the use of the orthogonality relations for characters of finite fields in order to evaluate explicitly (or almost so), or estimate, some particularly important exponential sums. Then, building on experience, Section 2.5 is a semi-philosophical discussion of the most common heuristics used to "guess" how exponential sums should behave.

2.1. Gauss sums

The following describes the general Gauss sums over a finite field.

Definition 2.1 (General Gauss sums). Let $\mathbf{F}_q$ be a finite field with q elements, ψ an additive character and χ a multiplicative character of $\mathbf{F}_q$. The Gauss sum associated to ψ and χ is defined to be

$$\tau(\chi,\psi)=\sum_{x\in\mathbf{F}_q^\times}\chi(x)\psi(x). \tag{2.1}$$

Remark 2.2. In particular, note that $\tau(\chi,1)=0$ for all non-trivial character χ, $\tau(1,\psi)=-1$ for all non-trivial character ψ (because of the "missing" term at $x=0$) and $\tau(1,1)=q-1$. If $\chi\neq 1$, we can replace the sum over $\mathbf{F}_q^\times$ to one over $\mathbf{F}_q$ since $\chi(0)=0$, but this is not permitted for the trivial character with our convention.

Example 2.3. Assume q is odd, and let $\chi=\chi_2$ be the non-trivial character of order 2 of $\mathbf{F}_q^\times$. If ψ is non-trivial, using (1.8), we find

$$\begin{aligned}\tau(\chi_2,\psi)&=\sum_{x\in\mathbf{F}_q}\chi_2(x)\psi(x)\\&=\sum_{x\in\mathbf{F}_q}\Bigl(\sum_{y^2=x}1-1\Bigr)\psi(x)=\sum_{x\in\mathbf{F}_q}\psi(x^2).\end{aligned}$$

Since, for $q=p$ prime, we have noted that $\chi_2(x)=\left(\frac{x}{p}\right)$, this shows that this Gauss sum is equal to the quadratic Gauss sum already introduced, if $\psi(x)=e(x/p)$.

It is not possible in general to compute $\tau(\chi,\psi)$, generalizing Theorem 3 for quadratic Gauss sums. However, the variation with respect to ψ is easy to understand: if ψ_0 is a fixed non-trivial character of $\mathbf{F}_q$, and ψ is another non-trivial character, we can find $a\in\mathbf{F}_q^\times$ such that $\psi(x)=\psi_0(ax)$ for all x (Proposition 1.13), and we then have

$$\tau(\chi,\psi)=\overline{\chi(a)}\tau(\chi,\psi_0) \tag{2.2}$$

since

$$\sum_{x\in\mathbf{F}_q^\times}\chi(x)\psi_0(ax)=\sum_{y\in\mathbf{F}_q^\times}\chi(a^{-1}y)\psi_0(y)=\overline{\chi(a)}\tau(\chi,\psi_0).$$

Equally importantly, the *modulus* of Gauss sums is known:

Proposition 2.4. *Let $\mathbf{F}_q$ be a finite field with q elements, ψ a non-trivial additive character and χ a non-trivial multiplicative character of $\mathbf{F}_q$. Then we have*

$$|\tau(\chi,\psi)| = \sqrt{q}.$$

Proof. Let $\tau = \tau(\chi,\psi)$. Since the idea of getting a square root from a sum like this may seem strange, we try to compute $|\tau|^2$ instead. For this, we can expand the product of the Gauss sum and its conjugate, and doing so we get

$$\begin{aligned}|\tau|^2 &= \sum_{x,y\in\mathbf{F}_q^\times} \chi(x)\overline{\chi(y)}\psi(x)\overline{\psi(y)}\\ &= \sum_{x,y\in\mathbf{F}_q^\times} \chi(xy^{-1})\psi(x-y).\end{aligned}$$

We can write $u = xy^{-1}$, which ranges freely over $\mathbf{F}_q^\times$ for all fixed y, and we can therefore rearrange the sum as

$$|\tau|^2 = \sum_{u\in\mathbf{F}_q^\times} \chi(u) \sum_{y\in\mathbf{F}_q^\times} \psi(y(u-1))$$

where we have now isolated a pure additive inner sum. From the description of all additive characters of $\mathbf{F}_q$, this inner sum is of the type

$$\sum_{\alpha} \alpha(u-1) - 1 = \begin{cases} -1 & \text{if } u \neq 1\\ q-1 & \text{if } u = 1,\end{cases}$$

(by orthogonality, α running over all additive characters). Hence we obtain

$$|\tau|^2 = -\sum_{u\in\mathbf{F}_q^\times} \chi(u) + q = q,$$

by orthogonality again. □

This property of Gauss sums is quite remarkable: indeed, it is an algebraic integer (as a sum of roots of unity!) of modulus exactly $\sqrt{q}$. In fact, a further property holds: for any field automorphism σ of $\mathbf{C}$, we have

$$\sigma(\tau(\chi,\psi)) = \tau(\sigma\circ\chi, \sigma\circ\psi),$$

because, by general algebra, the composites $\sigma\circ\chi$ and $\sigma\circ\psi$ are, respectively, non-trivial multiplicative and additive characters of $\mathbf{F}_q$. So the proposition shows that Gauss sums give examples of *Weil numbers*:

Definition 2.5 (Weil number). Let q be a power of a prime and $m \in \mathbf{Z}$ an integer. A q-Weil number of weight m is an algebraic number α with either of the following equivalent properties:

(1) Any root $\beta \in \mathbf{C}$ of the minimal polynomial of α, including $\beta = \alpha$, is such that $|\beta| = q^{m/2}$.

(2) For any embedding $\iota : \mathbf{Q}(\alpha) \hookrightarrow \mathbf{C}$, we have $|\iota(\alpha)| = q^{m/2}$.

In this language, the Gauss sum $\tau(\chi,\psi)$ associated to non-trivial characters of $\mathbf{F}_q$ is a q-Weil number of weight 1. These, together with the even simpler roots of unity (q-Weil numbers of weight 0 for any q) are the simplest Weil numbers. However there are many others. For instance, let q be a power of a prime and let a be any integer with $|a| < 2\sqrt{q}$; the roots of the quadratic polynomial

$$X^2 - aX + q$$

are then q-Weil numbers of weight 1. Indeed, since the discriminant is < 0, the roots are complex conjugates of each other, say $(\alpha, \bar{\alpha})$, and therefore

$$|\alpha|^2 = \alpha\bar{\alpha} = q.$$

REMARK 2.6. From Gauss's result (Theorem 3), we know how to compute also the argument of quadratic Gauss sums. What about the argument of more complicated ones? It turns out that those arguments behave quite unpredictably in general. In fact, if we consider all the arguments $\theta_p(\chi) \in [0, 1[$ of the Gauss sums

$$\tau(\chi, \psi_0) = \sqrt{p}e(\theta_p(\chi))$$

for all non-trivial multiplicative characters χ of $\mathbf{F}_p^\times$, with the fixed additive character

$$\psi_0 \,:\, x \mapsto e\Bigl(\frac{x}{p}\Bigr),$$

we obtain a collection of $p - 2$ angles which become *equidistributed* in $[0, 1]$ as $p \to +\infty$. We recall the definition of this important concept, which is crucial to understanding the variation of exponential sums in families (this result is indeed a first case of Problem (4) of the introduction).

DEFINITION 2.7 (Equidistribution). Let X be a compact topological space and μ a Borel probability measure on X. Let (Y_n) be a sequence of non-empty finite sets with maps

$$\theta_n \,:\, Y_n \longrightarrow X.$$

Then the points $\{\theta_n(y)\}_{y \in Y_n}$ *become equidistributed with respect to* μ as $n \to +\infty$ if and only if, for any continuous function $f \,:\, X \to \mathbf{C}$, we have

$$\frac{1}{|Y_n|} \sum_{y \in Y_n} f(\theta_n(y)) \longrightarrow \int_X f(x) d\mu(x),$$

as $n \to +\infty$.

It is not too difficult to show that this is equivalent with

$$\frac{1}{|Y_n|} |\{y \in Y_n \mid \theta_n(y) \in U\}| \longrightarrow \mu(U),$$

for any open set $U \subset X$ with boundary ∂U of μ-measure 0.

We then have the following remarkable result of Deligne:

THEOREM 2.8 (Deligne). *As $p \to +\infty$, the angles $\{\theta_p(\chi)\}_{\chi \neq 1}$ in $[0, 1]$ become equidistributed with respect to Haar measure.*

The proof of this result is quite deep, as it involves the Riemann Hypothesis for exponential sums in an arbitrarily large number of variables (see, e.g. [12, §11.11] for a description of the proof, or the second part [14,] of this course). This illustrates the fact that issues concerning the variation of exponential sums in families are very deep, even for the Gauss sums which are among the simplest ones.

2.2. Kloosterman's bound

We can now describe the very nice proof of Theorem 9. First of all, we generalize the definition of Kloosterman sums as follows:

Definition 2.9 (Kloosterman sums over finite fields). Let $\mathbf{F}_q$ be a finite field with q elements, and let ψ, η be two additive characters of $\mathbf{F}_q$. The associated *Kloosterman sum* $S(\psi, \eta)$ is defined by

$$S(\psi,\eta)=\sum_{x\in\mathbf{F}_q^\times}\psi(x)\eta(x^{-1}). \tag{2.3}$$

For $q = p$, we can take the characters

$$\psi(x)=e(mx/p),\qquad \eta(x)=e(n^{-1}x/p),$$

for any m, $n \in \mathbf{F}_q^\times$, and then we we recover the sums of Definition 7:

$$S(m,n;p)=S(\psi,\eta).$$

As first easy remarks, we note that $S(\psi, \eta)$ is always a real number, since

$$\begin{aligned}\overline{S(\psi,\eta)}&=\sum_{x\in\mathbf{F}_q^\times}\overline{\psi(x)\eta(x^{-1})}=\sum_{x\in\mathbf{F}_q^\times}\psi(-x)\eta(-x^{-1})\\&=\sum_{y\in\mathbf{F}_q^\times}\psi(y)\eta(y^{-1})=S(\psi,\eta),\end{aligned}$$

by using the change of variable $y = -x$.

We also note the simple relation

$$S(\psi,\eta)=S(\psi_b,\eta_{b^{-1}}) \tag{2.4}$$

for any additive characters and $b \in \mathbf{F}_q^\times$, where $\psi_b(x) = \psi(bx)$ and $\eta_{b^{-1}}(x) = \eta(b^{-1}x)$ (by the bijective change of variable $x = by$ in the sum).

Kloosterman's idea to estimate Kloosterman sums is based on trying to understand all of them globally, and not individually. More precisely, the idea is to use the following fact: if we can prove an average bound[1]

$$\sum_{\psi,\eta}{}^{*}|S(\psi,\eta)|^{2k}\leqslant M \tag{2.5}$$

for some $k \geqslant 1$ and $M \geqslant 0$, then for any fixed ψ_0, η_0, we can deduce from (2.4) that

$$(q-1)|S(\psi,\eta)|^{2k}=\sum_{b\neq 1}|S(\psi_b,\eta_{b^{-1}})|^{2k}\leqslant M, \tag{2.6}$$

and hence

$$|S(\psi,\eta)|\leqslant\Bigl(\frac{M}{q-1}\Bigr)^{1/(2k)},$$

for every fixed ψ and η.

The left-hand side of (2.5) is called the *k-th moment of* $|S(\psi, \eta)|^2$, and is a well-known quantity from the probabilistic point of view, which suggests that it is natural to study it if we feel that the mapping

$$(\psi,\eta)\mapsto S(\psi,\eta)$$

associated to Kloosterman sums is quite "random" (which is a fairly reasonable thing to say).

Kloosterman proved the following formulas for the moments of small order:

[1] Using the shorthand notation $\sum_{\psi}^{*}$ to indicate a sum restricted to non-trivial additive characters.

Proposition 2.10. *Let $\mathbf{F}_q$ be a finite field with q elements, and for $k \geqslant 0$, let*

$$M_k = M_{k,q} = \frac{1}{(q-1)^2} \sideset{}{^*}\sum_{\psi,\eta} |S(\psi,\eta)|^{2k}.$$

Then we have

$$M_0 = 1, \qquad M_1 = \frac{q^2-q-1}{q-1}, \qquad M_2 = \frac{2q^3-3q^2-3q-1}{q-1}.$$

From the last formula, we immediately derive, using (2.6), that

$$|S(\psi,\eta)| \leqslant ((q-1)M_2)^{1/4} < 2q^{3/4}$$

for any pair of non-trivial characters, which – in particular – proves Theorem 9. Morever, we can also write the inequality

$$M_2 \leqslant \Bigl(\max_{\psi,\eta} |S(\psi,\eta)|^2\Bigr) \times M_1,$$

and thefore there exists some pair (ψ,η) of non-trivial characters for which

$$|S(\psi,\eta)|^2 \geqslant \frac{M_2}{M_1} = \frac{2q^3-3q^2-3q-1}{q^2-q-1} > 2q-2,$$

which gives a generalization of (7). (Using (2.4), one can even assume that ψ be fixed).

Proof. The case of M_0 is of course trivial, and to prove Proposition 2.10 in the other cases, we start with a general formula: for any $k \geqslant 1$, we claim that

$$M_k = \frac{q^2}{(q-1)^2}|A_k(\mathbf{F}_q)| - 2(q-1)^{-1} - (q-1)^{2k-2}, \tag{2.7}$$

where

$$A_k(\mathbf{F}_q) = \Bigl\{(x,y) \in (\mathbf{F}_q^\times)^{k+k} \mid \sum_{1\leqslant i\leqslant k} x_i = \sum_{1\leqslant i\leqslant k} y_i, \text{ and } \sum_{1\leqslant i\leqslant k} x_i^{-1} = \sum_{1\leqslant i\leqslant k} y_i^{-1}\Bigr\}. \tag{2.8}$$

This reduces the proof of the proposition to a *counting problem* over $\mathbf{F}_q$: we need to know the number of solutions of certain polynomial equations (possibly in many variables, if k is large) over $\mathbf{F}_q$.

To prove (2.7) is not difficult: we first add the contributions of the trivial characters to be able to apply orthogonality more efficiently, and subtract them[2] using the simple formulas

$$S(1,\eta) = \sum_{x\in\mathbf{F}_q^\times} \eta(x^{-1}) = \sum_{y\in\mathbf{F}_q^\times} \eta(y) = -1,$$

if $\eta \neq 1$, the analogue $S(\psi,1) = -1$ if $\psi \neq 1$, and

$$S(1,1) = q-1.$$

We have then

$$(q-1)^2 M_k = \sum_{\psi,\eta} |S(\psi,\eta)|^{2k} - 2(q-1) - (q-1)^{2k}.$$

[2] It may be instructive to check that, if one tries to do this at the next stage of the computation, there appear some serious complications.

Then by expanding the definition of the Kloosterman sums and their conjugates, we obtain that

$$(q-1)^2 M_k = \sum_{\psi,\eta} \sum_{\substack{x=(x_1,\ldots,x_k)\in(\mathbf{F}_q^\times)^k\\ y=(y_1,\ldots,y_k)\in(\mathbf{F}_q^\times)^k}} \cdots \sum \psi(x_1+\cdots+x_k-y_1-\cdots-y_k)$$
$$\times\, \eta(x_1^{-1}+\cdots+x_k^{-1}-y_1^{-1}-\cdots-y_k^{-1}) - 2(q-1)-(q-1)^{2k},$$

and appealing to orthogonality of additive characters, we deduce that

$$\begin{aligned}(q-1)^2 M_k &= \sum_{x,y}\Bigl(\sum_{\psi}\psi(T(x)-T(y))\Bigr)\\ &\quad\times\Bigl(\sum_{\eta}\eta(U(x)-U(y))\Bigr) - 2(q-1)-(q-1)^{2k}\\ &= q^2|A_k(\mathbf{F}_q)| - 2(q-1)-(q-1)^{2k},\end{aligned}$$

where we have writen $T(x)=x_1+\cdots+x_k$, $U(x)=x_1^{-1}+\cdots+x_k^{-1}$ for $x\in(\mathbf{F}_q^\times)^k$. This gives (2.8).

We must now compute $|A_1(\mathbf{F}_q)|$ and $|A_2(\mathbf{F}_q)|$. For the former, we have

$$\begin{aligned}A_1(\mathbf{F}_q) &= \{(x,y)\in\mathbf{F}_q^\times\times\mathbf{F}_q^\times \mid x=y,\ x^{-1}=y^{-1}\}\\ &= \{(x,x)\mid x\in\mathbf{F}_q^\times\}\end{aligned}$$

so that $|A_1(\mathbf{F}_q)|=q-1$. Thus

$$(q-1)^2 M_1 = q^2(q-1)-2(q-1)-(q-1)^2,$$

so that

$$M_1 = \frac{q^2-q-1}{q-1}.$$

The case of $k=2$ requires a bit more care. The equations to solve are now given by

$$\begin{cases} x_1+x_2 = y_1+y_2\\ \dfrac{1}{x_1}+\dfrac{1}{x_2}=\dfrac{1}{y_1}+\dfrac{1}{y_2}.\end{cases}$$

There are obvious solutions, given by taking $y=(y_1,y_2)$ to be a permutation of (x_1,x_2):

$$(x,y)=(x_1,x_2,x_1,x_2), \quad\text{or}\quad (x_1,x_2,x_2,x_1)\ ;$$

taking account repetition, there are $2(q-1)^2-(q-1)$ such solutions.

Having found these, we are tempted to try to find the (possible) others by attempting to see under which conditions the quantities $(x_1+x_2, x_1^{-1}+x_2^{-1})$ determine the pair $(x_1,x_2)\in(\mathbf{F}_q^\times)^2$ up to permutation – for values of this type, the only solutions corresponding to (x_1,x_2) will precisely be the ones above.

Now, from the theory of symmetric functions, we know that the pair up to permutation is determined exactly by the elementary symmetric functions (x_1+x_2, x_1x_2). We already have the first function, and for the other, we simply observe that

$$x_1^{-1}+x_2^{-1}=\frac{x_1+x_2}{x_1x_2},$$

and therefore we can indeed recover (x_1+x_2,x_1x_2) *provided* we have

$$x_1^{-1}+x_2^{-1}\neq 0,$$

which translates to $x_1 + x_2 \neq 0$. And indeed, we have another family of solutions given by

$$(x_1, -x_1, y_1, -y_1), \qquad (x_1, y_1) \in (\mathbf{F}_q^\times)^2.$$

There are $(q-1)^2$ of these but among these, the $2(q-1)$ solutions given by

$$(x_1, -x_1, x_1, -x_1), \qquad (x_1, -x_1, -x_1, x_1)$$

have already been counted above. Therefore, we derive

$$\begin{aligned} |A_2(\mathbf{F}_q)| &= 2(q-1)^2 - (q-1) + (q-1)^2 - 2(q-1) \\ &= 3(q-2)(q-1), \end{aligned}$$

and from this, we get

$$M_2 = \frac{2q^3 - 3q^2 - 3q - 1}{q-1},$$

as we had claimed! □

REMARK 2.11. As already mentioned in the previous chapter, the result proved by Kloosterman is not best possible. We will prove later that

$$|S(\psi, \eta)| \leqslant 2\sqrt{q}$$

for any non-trivial characters ψ and η over $\mathbf{F}_q$, generalizing Weil's bound (Theorem 10). In fact, this will follow immediately from the possibility of writing

$$S(\psi, \eta) = \alpha(\psi, \eta) + \beta(\psi, \eta),$$

where $\alpha(\psi, \eta)$, $\beta(\psi, \eta)$ are q-Weil numbers of weight 1. This expression as a combination of two Weil numbers is the analogue for Kloosterman sums of the simpler fact that Gauss sums are q-Weil numbers.

2.3. Jacobi sums

Jacobi sums form another class of exponential sums which are of great importance in algebraic number theory. We introduce them partly for purely aesthetic reasons, and partly as a way to point out some interesting analogies with some classical functions defined by integrals...

DEFINITION 2.12 (Jacobi sums). Let $\mathbf{F}_q$ be a finite field with q elements, and let χ, ϕ be multiplicative characters of $\mathbf{F}_q$. The *Jacobi sum* associated to χ and ϕ is given by

$$J(\chi, \phi) = \sum_{x \in \mathbf{F}_q} \chi(x)\phi(1-x) = \sum_{x+y=1} \chi(x)\phi(y).$$

These sums turn out, rather surprisingly, to be expressible in terms of general Gauss sums.

PROPOSITION 2.13. *Let $\mathbf{F}_q$ be a finite field with q elements, let χ and ϕ be non-trivial multiplicative characters such that $\chi\phi$ is also non-trivial. Fix a non-trivial additive character ψ of $\mathbf{F}_q$. We then have*

$$J(\chi, \phi) = \frac{\tau(\chi, \psi)\tau(\phi, \psi)}{\tau(\chi\phi, \psi)}.$$

In particular, we have then

$$|J(\chi, \phi)| = \sqrt{q}.$$

Note that, by Proposition 2.4, the denominator in the formula is non-zero.

Proof. Once the formula for the Jacobi sum is established, we obtain its modulus immediately by applying Proposition 2.4 to the Gauss sums that occur.

Now, for the proof, it is natural to try to compute

$$J(\chi, \phi)\tau(\chi\phi, \psi)$$

since otherwise it seems difficult to envision how to perform a division by a Gauss sum.

Expanding the definitions of the two sums, we get

$$J(\chi, \phi)\tau(\chi\phi, \psi) = \sum_{x \in \mathbf{F}_q} \sum_{y \in \mathbf{F}_q^\times} \chi(x)\phi(1-x)\chi(y)\phi(y)\psi(y).$$

The sum may be restricted to $x \notin \{0, 1\}$, since χ and ϕ are non-trivial. Then we can define $u = xy$ and $v = y - xy$, and we obtain a bijective change of variable from $(x, y) \in (\mathbf{F}_q - \{0, 1\}) \times \mathbf{F}_q^\times$ to

$$\{(u, v) \in \mathbf{F}_q^\times \times \mathbf{F}_q^\times \mid u + v \neq 0\}$$

since we can recover x and y by $y = v + u$, $x = u/(u + v)$. We derive

$$\begin{aligned} J(\chi, \phi)\tau(\chi\phi, \psi) &= \sum\sum_{\substack{u,v \in \mathbf{F}_q^\times \\ u+v \neq 0}} \chi(u)\phi(v)\psi(u + v) \\ &= \tau(\chi, \psi)\tau(\phi, \psi) - \sum_{u \in \mathbf{F}_q^\times} \chi(u)\phi(-u) \\ &= \tau(\chi, \psi)\tau(\phi, \psi) \end{aligned}$$

since $\chi\phi$ is also non-trivial. □

Remark 2.14. It is immediate that if σ is a field automorphism of $\mathbf{C}$, we have

$$\sigma(J(\chi, \phi)) = J(\sigma \circ \chi, \sigma \circ \phi),$$

and therefore the proposition also proves that $J(\chi, \phi)$, under the conditions there, is a q-Weil number of weight 1 (see Definition 2.5).

Despite appearances, the Jacobi sums have rather different properties than Gauss sums, coming from their definition in terms of multiplicative characters. For instance, we derive quickly a proof of a well-known theorem of Fermat:

Theorem 2.15 (Fermat). *Let p be a prime number such that $p \equiv 1 \pmod 4$. Then there are integers a, b such that*

$$p = a^2 + b^2.$$

Proof. Because $p \equiv 1 \pmod 4$, there exists a character χ of order 4 of $\mathbf{F}_p^\times$. Let χ_2 denote the Legendre character of order 2, and consider the Jacobi sum $J = J(\chi, \chi_2)$. By Proposition 2.13, we have then $|J|^2 = p$. But J is a sum of terms of the type

$$\chi(x)\chi_2(y),$$

and $\chi_2(y) \in \{-1, 0, 1\}$, while $\chi(x) \in \{-i, -1, 0, 1, i\}$, since $\chi(x)^4 = 1$. Thus J can be written $J = a + bi$ with a, $b \in \mathbf{Z}$. Then

$$p = |J|^2 = a^2 + b^2.$$

□

Remark 2.16. The appearance of the formula for Jacobi sums may remind the reader of a well-known formula of Euler for his beta function:

$$B(a,b)=\int_0^1 x^a(1-x)^b dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

were the gamma function

$$\Gamma(s)=\int_0^{+\infty} e^{-x}x^s\frac{dx}{x} \tag{2.9}$$

plays the role of the Gauss sums. This analogy is not fortuitous: note that $x\mapsto e^{-x}$ is a character of the additive group $\mathbf{R}$, while $x\mapsto x^s$, for $s\in\mathbf{C}$, is a character of the multiplicative group $(\mathbf{R}^+)^\times$, and the measure $x^{-1}dx$ has the property of being invariant under (multiplicative) translation

$$\int_0^{+\infty} f(ax)\frac{dx}{x}=\int_0^{+\infty} f(x)\frac{dx}{x}$$

for any $a>0$ (provided the integrals converge). Thus (2.9) is wholly similar to (2.1), and the definition of the Beta function also reflects this analogy. It turns out that most of the identities satisfied by the gamma function have analogues in the context of Gauss sums.

In particular, recall the duplication formula

$$\Gamma(s)\Gamma(s+\tfrac{1}{2})=\sqrt{\pi}2^{1-2s}\Gamma(2s)=2^{1-2s}\Gamma(\tfrac{1}{2})\Gamma(2s),$$

(usually attributed to Gauss and Legendre); its analogue is a relation between $\tau(\chi,\psi)$, $\tau(\chi\chi_2,\psi)$ and $\tau(\chi^2,\psi)$ (for q odd) where the character χ_2 of order 2 replaces the shift by $1/2$. Precisely, a special case of the so-called *Hasse-Davenport product relation* states:[3]

Proposition 2.17. *Let q be odd, let $\mathbf{F}_q$ be a field with q elements and χ_2 the non-trivial multiplicative character of order 2 of $\mathbf{F}_q$. Then, for any multiplicative character χ and any non-trivial additive character ψ of $\mathbf{F}_q$, we have*

$$\tau(\chi^2,\psi)\tau(\chi_2,\psi)=\chi(4)\tau(\chi,\psi)\tau(\chi\chi_2,\psi), \tag{2.10}$$

Proof. This is quite simple using the relations between Gauss sums and Jacobi sums. First, note that it is a tautology for $\chi=1$ or $\chi=\chi_2$ (the latter because $\chi_2(4)=1$), and otherwise, we have

$$\frac{\tau(\chi,\psi)^2}{\tau(\chi^2,\psi)}=J(\chi,\chi)=\sum_{x\in\mathbf{F}_q}\chi(x(1-x)), \tag{2.11}$$

using (1.9) and Proposition 2.13. We now write

$$\begin{aligned}\sum_{x\in\mathbf{F}_q}\chi(x(1-x))&=\sum_{y\in\mathbf{F}_q}\chi(y)\sum_{x-x^2=y}1\\&=\sum_{y\in\mathbf{F}_q}\chi(y)\times(1+\chi_2(1-4y))\end{aligned}$$

since the quadratic equation $x-x^2=y$ has, in $\mathbf{F}_q$, the same number of solutions of the discriminant equation $\Delta^2=1-4y$, which is given by using (1.8) with $d=2$ (recall q is odd here).

[3] There is another formula known as the Hasse-Davenport relation, which we will consider in the next chapter.

Expanding and using the fact that $\chi \neq 1$, we find

$$J(\chi,\chi) = \sum_{y\in\mathbf{F}_q} \chi(y)\chi_2(1-4y) = \overline{\chi(4)}J(\chi,\chi_2) = \overline{\chi(4)}\frac{\tau(\chi,\psi)\tau(\chi_2,\psi)}{\tau(\chi\chi_2,\psi)}$$

by Proposition 2.13 again. Comparing with (2.11), we see that (2.10) is proved. □

2.4. Salié sums

Our last example is meant to illustrate once more some of the points already presented up to now. It is also a very important type of sums, playing a crucial role in a number of deep arithmetic problems.

Definition 2.18 (Salié sums). Let $\mathbf{F}_q$ be a finite field with q elements, with q odd. Let χ_2 denote the unique non-trivial multiplicative character of order 2 of $\mathbf{F}_q^\times$, and let ψ, η be additive characters of $\mathbf{F}_q$. The associated *Salié sum* is defined by

$$T(\psi,\eta) = \sum_{x\in\mathbf{F}_q^\times} \chi_2(x)\psi(x)\eta(x^{-1}).$$

This looks much like the Kloosterman sum (2.3), and one may expect similar properties. This is the case, with one major surprise: the Salié sums turns out to be expressible elementarily and (almost) explicitly as a sum of two q-Weil numbers of weight 1!

Theorem 2.19 (Salié). *Let $\mathbf{F}_q$ be a finite field with q elements, q odd, and let ψ, η be non-trivial additive characters of $\mathbf{F}_q$. Then we have*

$$T(\psi,\eta) = \tau(\chi_2,\psi) \sum_{y^2=4a} \psi(y), \tag{2.12}$$

where $a \in \mathbf{F}_q^\times$ is such that $\eta(x) = \psi(ax)$ for all $x \in \mathbf{F}_q$.

In particular, $T(\psi,\eta)$ is a sum of two q-Weil numbers of weight 1, and we have

$$|T(\psi,\eta)| \leqslant 2\sqrt{q}. \tag{2.13}$$

Proof. The idea is to study the variation with respect to $b \in \mathbf{F}_q^\times$ of

$$\varphi(b) = T(\psi_b,\eta) = \sum_{x\in\mathbf{F}_q^\times} \chi_2(x)\psi(bx + ax^{-1}).$$

More precisely, we endeavor to represent this function by a discrete multiplicative Fourier expansion, namely by Proposition 1.10 applied to $\mathbf{F}_q^\times$, we have

$$\varphi(b) = \sum_{\chi} \hat{\varphi}(\chi)\chi(b), \tag{2.14}$$

where χ runs over all multiplicative characters of $\mathbf{F}_q^\times$ and

$$\hat{\varphi}(\chi) = \frac{1}{q-1}\sum_{b\in\mathbf{F}_q^\times} \varphi(b)\overline{\chi(b)}.$$

We now compute the Fourier coefficients: by the definition of Salié sums, we have

$$\begin{aligned}\hat{\varphi}(\chi) &= \frac{1}{q-1}\sum_{b\in\mathbf{F}_q^\times}\varphi(b)\overline{\chi(b)} = \frac{1}{q-1}\sum_{b\in\mathbf{F}_q^\times}\overline{\chi(b)}\sum_{x\in\mathbf{F}_q^\times}\chi_2(x)\psi\Bigl(bx+\frac{a}{x}\Bigr)\\ &= \frac{1}{q-1}\sum_{x\in\mathbf{F}_q^\times}\chi_2(x)\psi(ax^{-1})\sum_{b\in\mathbf{F}_q^\times}\overline{\chi(b)}\psi(bx)\\ &= \frac{\tau(\bar{\chi},\psi)}{q-1}\sum_{x\in\mathbf{F}_q^\times}\chi_2(x)\chi(x)\psi(ax^{-1})\end{aligned}$$

so that

$$\hat{\varphi}(\chi) = \frac{\chi(a)\chi_2(a)\tau(\bar{\chi},\psi)\tau(\bar{\chi}\chi_2,\psi)}{q-1}, \tag{2.15}$$

after applying (2.2) twice. We see that we can now appeal to the Hasse-Davenport formula (2.10) – which is valid for all characters – to derive further that

$$\hat{\varphi}(\chi) = \frac{\chi_2(a)\tau(\chi_2,\psi)}{q-1}\chi(4a)\tau(\bar{\chi}^2,\psi) \tag{2.16}$$

(observe how, if the "twisting" factor χ_2 had been absent, or had been replaced with another character, we would not have been able to proceed with this step). And finally, using the expansion (2.14), we get

$$\begin{aligned}T(\psi,\eta) = \varphi(1) &= \sum_{\chi}\hat{\varphi}(\chi)\\ &= \frac{\chi_2(a)\tau(\chi_2,\psi)}{q-1}\sum_{\chi}\chi(4a)\tau(\bar{\chi}^2,\psi)\\ &= \frac{\chi_2(a)\tau(\chi_2,\psi)}{q-1}\sum_{\chi}\sum_{x\in\mathbf{F}_q}\chi(4ax^{-2})\psi(x)\\ &= \frac{\chi_2(a)\tau(\chi_2,\psi)}{q-1}\sum_{x\in\mathbf{F}_q}\psi(x)\sum_{\chi}\chi(4ax^{-2})\\ &= \chi_2(a)\tau(\chi_2,\psi)\sum_{y^2=4a}\psi(y),\end{aligned}$$

by orthogonality of the multiplicative characters. We can remove the factor $\chi_2(a)$, because if it is -1, then a is not a square in $\mathbf{F}_q$, and the inner sum is zero anyway.

Since $a \neq 0$, the equation $y^2 = 4a$ has, in $\mathbf{F}_q$, either 0 or 2 solutions. In the first case, of course, $T(\psi,\eta) = 0$ (and can be written as $\sqrt{q}-\sqrt{q}$ as a combination of Weil numbers); otherwise, if y is one solution, we get $T(\psi,\eta)$ has a sum of two terms, each of which is a root of unity multiplied with the q-Weil number $\tau(\chi_2,\psi)$, hence is itself a q-Weil number of weight 1. The bound (2.13) is then, of course, clear. □

REMARK 2.20. This is not the simplest proof, which is probably one due to P. Sarnak, which is reproduced for instance in [12, Lemma 12.4], and where the idea is to analyze by additive Fourier transform the function

$$\tilde{\varphi}(b) = T(\psi_{b^2},\eta),$$

which can be done without appealing to the Hasse-Davenport relation or anything more involved than the orthogonality relations for characters. In fact, using this proof, and comparing (2.15) with the expansion in multiplicative characters of the resulting formula for our function $\varphi(b)$, one would derive another proof of (2.10).

Remark 2.21. An analogy quite similar to the one that we pointed out between Gauss sums and the gamma function (Remark 2.16) may help understand why the Salié sum could be simpler to handle than the Kloosterman sum.

Indeed, natural analogues of the more general exponential sums given by

$$S_\chi(\psi,\eta)=\sum_{x\in\mathbf{F}_q^\times}\chi(x)\psi(x)\eta(x^{-1})$$

are the K-*Bessel functions* defined by the integrals

$$K_\nu(x)=\frac{1}{2}\int_0^{+\infty}t^{-\nu}\exp\Bigl(-\frac{x}{2}\Bigl(t+\frac{1}{t}\Bigr)\Bigr)\frac{dt}{t},$$

where we recognize, as in Remark 2.16, the exponentials replacing additive characters and power functions replacing the multiplicative ones, and the invariant measure $t^{-1}dt$ (the normalization and the choice of $t^{-\nu}$ instead of t^ν is due to historical reasons). The K-Bessel functions satisfy the differential equation

$$x^2y''+xy'-(x^2+\nu^2)y=0$$

(and can be characterized among those solutions as the unique one such that $K_\nu(x)\sim(\frac{\pi}{2x})^{-1/2}e^{-x}$ as $x\to+\infty$).

One may think of K_0 or K_1 as an analogue of the standard Kloosterman sums, and of $K_{1/2}$ as an analogue of Salié sums; corresponding to the formula (2.12) is the fact that $K_{1/2}$ (and more generally all functions $K_{n+1/2}$, where $n\geqslant 0$ is an integer) is an elementary function, whereas K_0 and K_1 are not: we have indeed

$$K_{1/2}(x)=\Bigl(\frac{\pi}{2x}\Bigr)^{1/2}e^{-x}\tag{2.17}$$

(and one can show using differential Galois theory that K_1 can not be expressed as a finite combination of elementary functions). The formula

$$J_{1/2}(x)=\Bigl(\frac{\pi}{2x}\Bigr)^{1/2}\sin(x)=\Bigl(\frac{\pi}{2x}\Bigr)^{1/2}\frac{e^{ix}-e^{-ix}}{2},$$

is even closer in appearance to (2.12); it applies to the J-Bessel function defined by

$$J_\nu(x)=\frac{1}{2i\pi}\int_C t^{-\nu}\exp\Bigl(\frac{x}{2}\Bigl(t-\frac{1}{t}\Bigr)\Bigr)\frac{dt}{t}$$

where C is an arbitrary contour in the complex plane enclosing once the origin (in counterclockwise direction), e.g., the unit circle.

In this respect, our computation leading to (2.16) corresponds, roughly, to the following fact: the Mellin transform of the exponential appearing in (2.17) is given by the gamma function. The idea of the proof itself can be guessed if one knows[4] the Mellin transform of a K-Bessel function, namely

$$\int_0^{+\infty}K_\nu(x)x^s\frac{dx}{x}=2^{s-2}\Gamma\Bigl(\frac{s+\nu}{2}\Bigr)\Gamma\Bigl(\frac{s-\nu}{2}\Bigr)$$

a product of two gamma functions, which corresponds to (2.15). If $\nu=1/2$ (and only then), this is of the form $\Gamma(u)\Gamma(u+1/2)$ (for $u=s/2-1/4$), in which case one can apply

[4] This is a very classical fact, but of course Bessel functions are not necessarily part of the standard curriculum nowadays.

the duplication formula to reduce it to a single gamma factor $\Gamma(2u) = \Gamma(s - 1/2)$ – as we did with the Hasse-Davenport formula. The inverse Mellin transform of this leads to the formula for $K_{1/2}(x)$.

The best understanding of this type of analogies depends on the deeper analysis of variations of Kloosterman sums and Salié sums; from there, it results that there is some kind of "Galois group" attached to this variation, which is also analogous to the differential Galois group associated to the differential equations solved by Bessel functions. It turns out that these Galois groups are *solvable* in the case of Salié sums and Bessel functions of half-integral order, but non-solvable in the case of Kloosterman sums or generic Bessel functions. Being solvable, in the Bessel case, means exactly that the latter can be representated elementarily.

2.5. What to expect when you're estimating?

The computations in this section (Gauss sums, Jacobi sums and Salié sums particularly) have the following common feature: an exponential sum of oscillatory nature involving roughly q terms, turns out to be of modulus equal to, or bounded by, $\sqrt{q}$, up to some multiplicative constant. This is in fact a very general principle, which can be properly understood using probabilistic heuristics: a "random" sum of n complex numbers of modulus 1 is, unless very special circumstances apply, usually of size roughly $\sqrt{n}$. A convincing rigorous formulation can be given in probabilistic language using the fundamental limit theorem (also called "central" limit theorem) of probability: let (X_n) be a sequence of independent random variables, defined on some probability space $(\Omega, \Sigma, \mathbf{P})$, each of which takes complex values which are uniformly distributed on the unit circle (in other words,

$$X_n = e(U_n),$$

where the (U_n) are also independent, and are uniformly distributed on $[0, 1]$). Then we can form the sums

$$S_N = X_1 + \cdots + X_N = \sum_{n \leqslant N} e(U_n), \tag{2.18}$$

which are models of random exponential sums of length N. We ask about the size of these sums, and one finds that they are typically of size $\sqrt{N}$, in the sense that $N^{-1/2} S_N$ has a well-defined limiting distribution:

Proposition 2.22 (Fundamental limit theorem). *With notation as above, for any fixed real numbers $a < b$, we have*

$$\lim_{N \to +\infty} \mathbf{P}\Bigl(a < \frac{S_N}{\sqrt{N}} < b\Bigr) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt.$$

Note that this suggests not only that sums of length N should be of size at most N, but also indicates that they can not be of significantly smaller size. This may be a very useful heuristic aid when trying to devise a strategy to prove certain statements: if one finds that a certain course of action requires *better than "square-root cancellation"*, then it must be examined very critically. It is not inconceivable that the sums involved may be of such special nature that something like this holds, but there has to be special features involved.

Of course, the exponential sums we consider form a very special subset of the general random sums (2.18). But one can easily find many ways to confirm the heuristic principle for more reasonable families. For instance, suppose that p is a fixed prime and look at the set Ω_p of all functions

$$f\,:\,\mathbf{F}_p\to\mathbf{F}_p,$$

and the corresponding sums

$$S(f)=\sum_{x\in\mathbf{F}_p}e\Bigl(\frac{f(x)}{p}\Bigr)$$

which we view as "random" exponential sums over the field $\mathbf{F}_p$ by looking at all possibilities for f, uniformly chosen. Then we claim that

$$\frac{1}{|\Omega_p|}\sum_{f\in\Omega_p}|S(f)|^2=p,$$

(which again confirms that, this time in mean-square average, the sums $S(f)$ are typically of size $\sqrt{p}$). To prove this, expand the modulus square on the left-hand side using the definition, and exchange the sum over f, to get

$$\frac{1}{|\Omega_p|}\sum_{f\in\Omega_p}|S(f)|^2=\sum_{x,y\in\mathbf{F}_p}\frac{1}{|\Omega_p|}\sum_{f\in\Omega_p}e\Bigl(\frac{f(x)-f(y)}{p}\Bigr).$$

Now we claim that the inner sum is zero unless $x=y$ (the "diagonal terms"), in which case it is equal to 1. Clearly this implies the mean-square statement.

To see the claim, note that it is obvious if $x=y$; otherwise, for any $a\in\mathbf{F}_p$, the number of $f\in\Omega_p$ with

$$f(x)-f(y)=a$$

is the same (i.e., it is $|\Omega_p|/p$); indeed, the map

$$f\mapsto f(x)-f(y)$$

can be viewed as a linear form on the $\mathbf{F}_p$-vector space $\Omega_p\simeq\mathbf{F}_p^p$, and if it is non-zero (which is the case when $x\neq y$, as one can take for f the characteristic function of the singleton $\{x\}$ which maps to 1), it is surjective and each equation $f(x)-f(y)=a$ defines a hyperplane in Ω_p.

Consequently

$$\frac{1}{|\Omega_p|}\sum_{f\in\Omega_p}e\Bigl(\frac{f(x)-f(y)}{p}\Bigr)=\frac{1}{p}\sum_{a\in\mathbf{F}_p}e\Bigl(\frac{a}{p}\Bigr)=0$$

by orthogonality.

CHAPTER 3

The Riemann hypothesis for sums in one variable: introduction

In this brief chapter, we introduce briefly one of the two main topics of this book: a proof of the Riemann Hypothesis for exponential sums in one variable over a finite field – one of the finest achievements of A. Weil. The actual proofs are found in the next two chapters: first in the case of multiplicative character sums, then in the case of additive characters, where we also consider mixed sums.

3.1. What is the Riemann Hypothesis?

We start with an introductory section that will both state the main theorems to be proved in this chapter, and explain *why* those results are analogues of the classical Riemann Hypothesis for the Riemann zeta function, and for Dirichlet L-functions. This will provide helpful motivation for the strategy of the proof, which is described afterwards (even though the Riemann Hypothesis is far from proved, it turns out that some of the techniques used to prove intermediate results in its direction – such as the Prime Number Theorem – are very relevant to the situation over finite fields).

We will prove the following two general results for multiplicative and additive character sums (an analogue result for "mixed" sums, for instance Salié, will appear as a sequence of exercises). We start with the multiplicative case:[1]

Theorem 3.1 (A. Weil; Riemann Hypothesis for one-variable multiplicative sums). *Let $\mathbf{F}_q$ be a finite field with q elements. Fix a polynomial $g \in \mathbf{F}_q[X]$, an integer $d \mid q-1$, $d \neq 1$, and a multiplicative character χ of $\mathbf{F}_q$ of order d, in particular non-trivial.*

Let m, $1 \leqslant m \leqslant \deg(g)$ be the number of distinct zeros of g in $\bar{\mathbf{F}}_q$. Then, provided *there is no polynomial $h \in \bar{\mathbf{F}}_q[X]$ such that $g = h^d$, we have*

$$\Bigl|\sum_{x \in \mathbf{F}_q} \chi(g(x))\Bigr| \leqslant (m-1)\sqrt{q}. \tag{3.1}$$

The additive case is as follows:

Theorem 3.2 (A. Weil; Riemann Hypothesis for one-variable additive sums). *Let $\mathbf{F}_q$ be a finite field with q elements. Fix a polynomial $f \in \mathbf{F}_q[X]$ of degree d and a non-trivial additive character ψ of $\mathbf{F}_q$.*

Then, provided

$$d < q \text{ and } (d, q) = 1, \tag{3.2}$$

we have

$$\Bigl|\sum_{x \in \mathbf{F}_q} \psi(f(x))\Bigr| \leqslant (d-1)\sqrt{q}. \tag{3.3}$$

[1] We will explain in detail how the multiplicative case is different, and in many sense easier, than the additive one.

Note that the condition $d < q$ in this last result is redundant: the estimate (3.3) is trivial if $d \geqslant q$; however the proof will provide a more general statement (over extension fields) where the condition is necessary.

Here are direct corollaries which illustrate the generality which is achieved; both are important results in their own right in applications to analytic number theory.

Corollary 3.3 (Hasse). *Let $g \in \mathbf{Z}[X]$ be a* cubic *polynomial with no multiple root in $\mathbf{C}$, or in other words, such that the discriminant of f is non-zero; for instance, take*

$$g = X^3 + aX + b$$

where $\Delta = a^3 - 27b^2 \neq 0$. Then we have

$$\Big| \sum_{0 \leqslant x \leqslant p-1} \Big(\frac{g(x)}{p}\Big) \Big| \ll \sqrt{p},$$

for all primes p, where the implied constant depends only on g.

Proof. Of course we want to apply Theorem 3.1 to $\mathbf{F}_q = \mathbf{Z}/p\mathbf{Z}$ and the reduction of g modulo p with the character χ given by the Legendre character. Now we claim that there exists a prime p_0 such that the theorem *does* apply for $p \geqslant p_0$ with $m = 3$. Indeed, χ is non-trivial as soon as $p \geqslant 3$. Moreover, since $\deg(g) = 3$, the reduction of g is itself of degree 3 for p not dividing the leading coefficient a_3 of g; as such, $g \pmod p$ is not a second power for such primes. Finally, since the discriminant Δ of g is non-zero in $\mathbf{Z}$, it follows that if p does not divide the discriminant of g, its reduction has also three distinct roots. Thus we have

$$\Big| \sum_{0 \leqslant x \leqslant p-1} \Big(\frac{g(x)}{p}\Big) \Big| \leqslant 2\sqrt{p},$$

for all primes $p \geqslant p_0 = \max(3, p_1, p_2)$, where p_1 is the largest prime divisor of a_3, and p_2 that of Δ.

We can now incorporate the "small" primes by a well-known trick:

$$\Big| \sum_{0 \leqslant x \leqslant p-1} \Big(\frac{g(x)}{p}\Big) \Big| \leqslant 2\sqrt{p_0}\sqrt{p}$$

for all primes. □

Corollary 3.4 (Weil). *Let $f \in \mathbf{Z}[X]$ be a non-constant polynomial of degree d. Then we have*

$$\Big| \sum_{0 \leqslant x \leqslant p-1} e\Big(\frac{f(x)}{p}\Big) \Big| \ll \sqrt{p},$$

for all primes p, where the implied constant depends only on f.

Proof. We apply the same trick as before with Theorem 3.2: for p large enough, the degree of f is also d, and we have (3.2) if $p > d$. Thus we get

$$\Big| \sum_{0 \leqslant x \leqslant p-1} e\Big(\frac{f(x)}{p}\Big) \Big| \leqslant (d-1)\sqrt{p}$$

for $p \geqslant p_0$ for some p_0, and obtain the final result by replacing the constant $d-1$ by $\max(d-1, \sqrt{p_0})$. □

We now recall the statement, and the arithmetic significance, of the classical Riemann Hypothesis. In fact, for reasons which we be clear very soon, we consider the *Generalized* Riemann Hypothesis for Dirichet L-functions. We won't give proofs here, but refer to standard textbooks, such as [12, §5].

Fix some integer $m \geqslant 1$; a *Dirichlet character* χ *modulo* m is a map

$$\mathbf{Z} \xrightarrow{\chi} \mathbf{C}$$

defined by

$$\chi(n) = \begin{cases} 0 & \text{if } (n,m) \neq 1 \\ \chi_r(n) & \text{if } (n,m) = 1, \end{cases}$$

for some multiplicative character

$$\chi_r \,:\, (\mathbf{Z}/m\mathbf{Z})^\times \to \mathbf{C}^\times.$$

Associated with χ is its L-function defined by

$$L(s,\chi) = \sum_{n \geqslant 1} \frac{\chi(n)}{n^s}$$

for all complex numbers with $\mathrm{Re}(s) > 1$ (since $|\chi(n)| \leqslant 1$ for all n, this is absolutely convergent in this region, and defines a holomorphic function there).

For the special case $m = 1$, $\chi(n) = 1$ for all n, we obtain the Riemann zeta function, denoted $\zeta(s)$. In general, these functions were introduced by Dirichlet to prove that there exist infinitely many primes in any arithmetic progression $p \equiv a \,(\mathrm{mod}\, m)$ if $(a,m) = 1$. Two of the main ingredients for that purpose were the *orthogonality relations* for characters modulo m, used to write

$$\sum_{\substack{n \geqslant 1 \\ n \equiv a \,(\mathrm{mod}\, m)}} \Lambda(n) n^{-s} = \frac{1}{\varphi(m)} \sum_{\chi \,(\mathrm{mod}\, m)} \overline{\chi(a)} \sum_{n \geqslant 1} \chi(n)\Lambda(n) n^{-s}$$

(for suitable weights $\Lambda(n)$ used to detect the primes), and the *Euler product expansion*

$$L(s,\chi) = \prod_p (1 - \chi(p) p^{-s})^{-1},$$

which leads to a link between the L-functions and the distribution of primes. This is most commonly expressed by taking the logarithmic derivative of this product, which – in the region of holomorphy – leads to

$$-\frac{L'}{L}(s,\chi) = \sum_{n \geqslant 1} \Lambda(n)\chi(n) n^{-s}$$

where the von Mangoldt function Λ is defined by

$$\Lambda(n) = \begin{cases} 0 & \text{if } n \text{ is not a power of a prime} \\ \log p & \text{if } n = p^m \text{ with } p \text{ prime, } m \geqslant 1 \end{cases}$$

(in particular, it detects primes, up to higher powers thereof). Combining with the orthogonality relation, we obtain

$$\sum_{\substack{n \geqslant 1 \\ n \equiv a \,(\mathrm{mod}\, m)}} \Lambda(n) n^{-s} = -\frac{1}{\varphi(m)} \sum_{x \,(\mathrm{mod}\, m)} \overline{\chi(a)}\, \frac{L'}{L}(s,\chi).$$

A further step in Dirichlet's proof is the fact that $L(s,\chi)$ has a meromorphic continuation to $\mathbf{C}$, and is in fact entire if $\chi_r \neq 1$. This continuation is particularly clear in the case when χ is *primitive*. This means that, in parallel to the prime factorization

$$m = \prod_{p|m} p^{v_p}$$

of m and the decomposition

$$(\mathbf{Z}/m\mathbf{Z})^{\times} \simeq \prod_{p|m} (\mathbf{Z}/p^{v_p}\mathbf{Z})^{\times}$$

given by the Chinese Remainder Theorem, χ_r factors as a product

$$\chi_r = \prod_{p|m} \chi_{r,p}$$

where each of the character $\chi_{r,p}$ of $(\mathbf{Z}/p^{v_p}\mathbf{Z})^{\times}$ is itself non-trivial. Then $L(s,\chi)$ satisfies a *functional equation*:

$$\Lambda(s,\chi) = i^{a(\chi)}\tau(\chi)m^{-s}\Lambda(1-s,\bar{\chi}), \tag{3.4}$$

where $a(\chi) = 0$ if $\chi(-1) = 1$, and $a(\chi) = 1$ otherwise, while

$$\Lambda(s,\chi) = \pi^{-s/2}\Gamma\Big(\frac{s+a(\chi)}{2}\Big)L(s,\chi),$$
$$\tau(\chi) = \sum_{x\in(\mathbf{Z}/m\mathbf{Z})^{\times}} \chi(x)e\Big(\frac{x}{m}\Big),$$

the latter being of course a Gauss sum for the finite ring $\mathbf{Z}/m\mathbf{Z}$. (This analytic function Λ should not be confused with the von Mangoldt function.)

Now the Generalized Riemann Hypothesis of $L(s,\chi)$ is the following conjectural statement: if χ is primitive, then any zero $\rho = \beta + i\gamma$ of $L(s,\chi)$ such that $0 < \beta < 1$ satisfies

$$\beta = \mathrm{Re}(\rho) = 1/2.$$

Taking the logarithmic derivative, we see that $-L'/L$ has then, in the critical strip

$$0 \leqslant \mathrm{Re}(s) \leqslant 1$$

only (i) simple poles at the zeros of $L(s,\chi)$, with residue giving by minus the multiplicity of the zeros; (ii) if χ is trivial, a simple pole with residue -1 at $s = 1$. The most immediate arithmetic significance of this is obtained by a well-known contour integration (already known to Riemann): it would follow that

$$\sum_{\substack{n\leqslant x\\ n\equiv a\,(\mathrm{mod}\,m)}} \Lambda(n) = \frac{x}{\varphi(q)} + O(x^{1/2}(\log qx)^2)$$

with an absolute implied constant.

In fact, using an easy summation by parts, one can check that the upper bounds

$$\sum_{\substack{n\leqslant x\\ n\equiv a\,(\mathrm{mod}\,m)}} \Lambda(n) = \frac{x}{\varphi(q)} + O(x^{\theta+\varepsilon})$$

for some $\theta \in [1/2, 1[$, all $\varepsilon > 0$ and $(a,m) = 1$ is *equivalent* with the assertion that all non-trivial zeros of Dirichlet L-functions $L(s,\chi)$ of modulus m satisfy $\mathrm{Re}(s) = 1/2$.

Because of the well-known analogy between integers and polynomials over a finite field, it seems natural to look at what the analogue of Dirichlet characters are in that context. This, it turns out, is one key tool in the study of exponential sums, although it works well only for purely multiplicative sums.

CHAPTER 4

Multiplicative character sums

4.1. Characters and L-functions for multiplicative exponential sums

Consider a finite field $\mathbf{F}_q$ with q elements and the ring $\mathbf{F}_q[X]$ of polynomials in one variable with coefficients in $\mathbf{F}_q$. It is a principal ideal domain, so to build an analogue of Dirichlet characters, it is natural fo fix a non-zero monic polynomial $g \in \mathbf{F}_q[X]$ and consider maps

$$\eta : \mathbf{F}_q[X] \to \mathbf{C}$$

defined by

$$\eta(f) = \begin{cases} 0 & \text{if } (f,g) \neq 1, \\ \eta_r(f \pmod g) & \text{if } (f,g) = 1, \end{cases}$$

(where the gcd is computed in $\mathbf{F}_q[X]$) and η_r is a group homomorphism

$$(\mathbf{F}_q[X]/g\mathbf{F}_q[X])^\times \to \mathbf{C}^\times.$$

Note that, as in the case of the usual Dirichlet characters, we have the multiplicativity relation

$$\eta(f_1 f_2) = \eta(f_1)\eta(f_2),$$

for *all* polynomials f_1, f_2.

As we will see, a more geometric language quickly becomes useful, based on the fact that elements of $\mathbf{F}_q[X]$ can be seen naturally as functions on $\mathbf{F}_q$ and its extensions (including $\bar{\mathbf{F}}_q$). The condition that f and g are coprime can be phrased in this language as saying that they do not have a common zero in $\bar{\mathbf{F}}_q$.

The link with exponential sums is given by the following simple lemma:

Proposition 4.1. *Let $\mathbf{F}_q$ be a finite field, $g \in \mathbf{F}_q[X]$ a non-constant monic polynomial. Given a non-trivial multiplicative character χ of $\mathbf{F}_q$, there exists a character η_r of $(\mathbf{F}_q[X]/g\mathbf{F}_q[X])^\times$ such that the corresponding Dirichlet character satisfies*

$$\eta(X - t) = \chi((-1)^{\deg(g)})\chi(g(t)),$$

for any $t \in \mathbf{F}_q$. In fact, there exists a unique such character η such that, for any $\nu \geqslant 1$ and $t \in \mathbf{F}_{q^\nu}$, we have

$$\chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(t))) = \chi(-1)^{\nu \deg(g)}\eta(\pi_t)^{\nu/d} \tag{4.1}$$

where π_t is the monic irreducible minimal polynomial of t, and $d = \deg \pi_t$ is its degree.

Thus the exponential sum

$$\sum_{x \in \mathbf{F}_q} \chi(g(x))$$

becomes inextricably linked with the Dirichlet character η modulo g.

Proof. We first factor

$$g = \prod_{\pi \mid g} \pi^{k_\pi}, \tag{4.2}$$

where $k_\pi \geqslant 1$ and the product runs over monic irreducible polynomials dividing g. The ideals generated by π^{k_π} are of course coprime, and hence by the Chinese Remainder Theorem, we have a group isomorphism

$$\begin{cases} (\mathbf{F}_q[X]/(g))^\times & \longrightarrow & \prod (\mathbf{F}_q[X]/(\pi^{k_\pi}))^\times \\ f & \mapsto & (f \pmod{\pi^{k_\pi}})_{\pi \mid g} \end{cases}. \tag{4.3}$$

Moreover, for all $\pi \mid g$, the quotient ring $\mathbf{F}_q[X]/(\pi)$ is a finite field since π is irreducible. Every polynomial, by euclidean division, has a unique representative of degree $\leqslant \nu_\pi = \deg(\pi)$, and hence this field is of order q^{ν_π}. Indeed, fixing arbitrarily some root $\alpha_\pi \in \bar{\mathbf{F}}_q$ of π, we have an isomorphism

$$\phi_\pi \begin{cases} \mathbf{F}_q[X]/(\pi) & \longrightarrow & \mathbf{F}_q(\alpha_\pi) \simeq \mathbf{F}_{q^{\nu_\pi}} \\ f & \mapsto & f(\alpha_\pi) \end{cases}. \tag{4.4}$$

Now any character η_r of $(\mathbf{F}_q[X]/(g))^\times$, by (4.3), can be described as the product of a tuple of characters $(\eta_{r,\pi})_{\pi \mid g}$ of $(\mathbf{F}_q[X]/(\pi^{k_\pi}))^\times$. Among these, we look at characters defined by

$$\eta_{r,\pi}(f) = \chi_\pi(\phi_\pi(f))^{k_\pi},$$

where χ_π is now a multiplicative character of the finite field

$$\mathbf{F}_q[X]/(\pi) \simeq \mathbf{F}_{q^{\nu_\pi}}.$$

Applying the character η constructed in this manner to $f = X - t$ we find first that

$$\eta(X - t) = 0$$

if $X - t$ is not coprime with g. That condition is equivalent with t being a zero of g, and in that case we also have

$$\chi(g(t)) = 0$$

by our usual convention. So we assume $(X - t, g) = 1$. In that case, following up on the definition of η using (4.3) and (4.4), we find

$$\begin{aligned} \eta(X - t) &= \prod_{\pi \mid g} \eta_{r,\pi}(X - t \pmod{\pi}) \\ &= \prod_{\pi \mid g} \chi_\pi(\phi_\pi(X - t))^{k_\pi} \\ &= \prod_{\pi \mid g} \chi_\pi((\alpha_\pi - t)^{k_\pi}). \end{aligned}$$

This is getting close to the target; in fact, if we remark that the choice of α_π leaves undesirable room for ambiguity, we are naturally led to assume that χ_π is of the type

$$\chi_\pi(x) = \tilde{\chi}_\pi(N_{\mathbf{F}_{q^{\nu_\pi}}/\mathbf{F}_q}(x)),$$

so that $\chi_\pi(x)$ depends only on the norm, and is therefore independent of the choice of an isomorphism $\mathbf{F}_q[X]/(\pi) \simeq \mathbf{F}_{q^{\nu_\pi}}$ (this is a consequence of the computation (1.3) of the kernel of the norm).

Using this, and the formula for the norm

$$N(\alpha_\pi - t) = \prod_{1 \leqslant k \leqslant \nu_\pi} (\alpha_\pi - t)^{q^k} = \prod_{1 \leqslant k \leqslant \nu_\pi} (\alpha_\pi^{q^k} - t)$$

for $t \in \mathbf{F}_q$, we are led to

$$\eta(X-t) = \prod_{\pi \mid g} \tilde{\chi}_\pi\Big(\prod_{1 \leqslant k \leqslant \nu_\pi} (\alpha_\pi^{q^k} - t)^{k_\pi}\Big).$$

And now we see that if $\tilde{\chi}_\pi = \chi$ for all $\pi \mid g$, we have exactly

$$\eta(X-t) = \chi((-1)^{\deg(g)})\chi(g(t)),$$

in view of the factorizations of g and of the irreducible monic polynomial π, namely

$$\pi = \prod_{1 \leqslant k \leqslant \nu_k} (X - \alpha_\pi^{q^k}).$$

There remains to prove the more general formula (4.1). For this, we first rewrite the definition based on our choice of η_r:

$$\eta(f) = \prod_{\pi \mid g} \chi(N_{\mathbf{F}_{q^{\nu_\pi}}/\mathbf{F}_q}(f(\alpha_\pi)))^{k_\pi}. \tag{4.5}$$

Let $\varpi = \pi_t$ be the minimal polynomial of $t \in \mathbf{F}_{q^\nu}$, $d = \deg(\varpi)$. We first note the factorizations

$$\varpi = \prod_{i=1}^{d} (X - t^{q^i}), \qquad \pi = \prod_{k=1}^{\nu_\pi} (X - \alpha_\pi^{q^k}) \quad \text{for } \pi \mid g,$$

and notice that the first one also gives

$$\varpi^{\nu/d} = \prod_{i=1}^{\nu} (X - t^{q^i}). \tag{4.6}$$

We start with $N_{k/\mathbf{F}_q} g(t)$ and develop it as a product using the factorizations of g and its factors π:

$$\begin{aligned} N_{k/\mathbf{F}_q} g(t) = \prod_{i=1}^{\nu} g(t^{q^i}) &= \prod_{\pi \mid g} \prod_{i=1}^{\nu} \pi(t^{q^i})^{k_\pi} \\ &= \prod_{\pi \mid g} \prod_{i=1}^{\nu} \prod_{k=1}^{\nu_\pi} (t^{q^i} - \alpha_\pi^{q^k})^{k_\pi}, \end{aligned}$$

and now we go backwards after exchanging the two innermost products and applying (4.6):

$$\begin{aligned} N_{k/\mathbf{F}_q} g(t) &= \prod_{\pi \mid g} \prod_{k=1}^{\nu_\pi} (-1)^{\nu k_\pi} \prod_{i=1}^{\nu} (\alpha_\pi^{q^k} - t^{q^i})^{k_\pi} \\ &= (-1)^{\nu \deg(g)} \Big(\prod_{\pi \mid g} N_{\mathbf{F}_{q^{\nu_\pi}}/\mathbf{F}_q}(\varpi(\alpha_\pi)^{k_\pi})\Big)^{\nu/d}, \end{aligned}$$

and the definition (4.5) gives precisely (4.1).

Finally the unicity of η follows because, specializing (4.1) to any root of any irreducible polynomial π of degree $\nu \geqslant 1$ (with $d = \nu$), we find that this formula determines the values of η for any monic irreducible polynomial. Using unique factorization and the multiplicativity of η, we can deduce uniquely the values of $\eta(f)$ for any monic f. But to deduce the general case, by multiplicativity, it is enough to know $\eta(\alpha)$ for $\alpha \in \mathbf{F}_q^\times$ a constant, which we can compute by selecting an arbitrary monic polynomial congruent to α modulo g (for instance $\alpha + g$ will do, since $\deg(g) \geqslant 1$). □

Remark 4.2. Using the recipe above, one can see that, in fact, we have

$$\eta(\alpha)=\prod_{\pi\mid g}\chi(\alpha^{\nu_\pi k_\pi})=\chi(\alpha)^{\deg(g)}, \tag{4.7}$$

for any $\alpha\in\mathbf{F}_q$.

We will say that η is the character *associated with the data* (g,χ) defining the multiplicative character sum.

Corollary 4.3. *Let $\mathbf{F}_q$ be a finite field, g and χ as above, and let η be the associated Dirichlet character.*

(1) *The character η is defined modulo the polynomial $g^\flat$ which is the product of the irreducible factors of g, without multiplicity.*

(2) *If χ is of order d and g is not a d-th power in $\bar{\mathbf{F}}[x]$, the associated Dirichlet character η is such that η_r is non-trivial.*

Proof. (1) is clear from the definition (4.5) that we used, since the irreducibles π appearing are exactly the irreducible factors of $g^\flat$.

(2) We use the notation in the proof above. If g is not a d-th power in $\bar{\mathbf{F}}_q[X]$, then for some $\pi_0\mid g$, of degree $\nu_0=\nu_\pi$, the multiplicity $k=k_{\pi_0}$, is not divisible by d, and therefore the character χ^k of $\mathbf{F}_q^\times$ is non-trivial. Hence, if $f\in\mathbf{F}_q[X]$ is such that

$$f(\alpha_\pi)=1 \text{ if } \pi\neq\pi_0,\qquad f(\alpha_{\pi_0})=\gamma\in\mathbf{F}_{q^{\nu_0}},$$

where

$$N_{\mathbf{F}_{q^{\nu_0}}/\mathbf{F}_q}(\gamma)=\beta,$$

for $\beta\neq 0$ such that $\chi(\beta)^k\neq 1$, we have $\eta(f)\neq 1$. Such a polynomial exists: first, γ exists, given β, by Lemma 1.3, and then the Chinese Remainder Theorem is applicable since the above argument allows us to rephrase the conditions in the form of congruences

$$f\equiv 1 \pmod{\pi},\ \pi\neq\pi_0,\qquad f\equiv p_\gamma \pmod{\pi_0},$$

where p_γ is a polynomial in $\mathbf{F}_q[X]$ with $p_\gamma(\alpha_{\pi_0})=\beta$ (given by the isomorphism (4.4)). □

Having obtained this, it is very natural to look at the L-functions of such characters. These turn out to be expressible in three different ways (one more than the Dirichlet series and Euler products of the usual L-functions).

Definition 4.4 (L-functions over finite fields). Let $\mathbf{F}_q$ be a finite field with q elements, g a non-zero monic polynomial in $\mathbf{F}_q[X]$, η a character of $\mathbf{F}_q[X]$ modulo g obtained from a multiplicative character η_r of $\mathbf{F}_q[X]/(g)$. The L-function attached to η is defined either as the Dirichlet series

$$L(\eta,s)=\sum_f \eta(f)|f|^{-s},$$

in its region of convergence, or as the formal power series

$$L(\eta,T)=\sum_f \eta(f)T^{\deg(f)},$$

where in both cases the variable f ranges over all non-zero monic polynomials in $\mathbf{F}_q[X]$.

Note that the two definitions are related by the formal substitution $T=q^{-s}$. We will now see that the series converges for $\mathrm{Re}(s)>1$, and has an analytic continuation to all of $\mathbf{C}$ as a meromorphic function. If η_r is non-trivial, the L-function is entire, and indeed this can be proved much more simply than the corresponding fact for the integers!

Proposition 4.5. *Let $\mathbf{F}_q$ be a finite field with q elements, g a non-zero monic polynomial in $\mathbf{F}_q[X]$, η a character of $\mathbf{F}_q[X]$ modulo g obtained from a multiplicative character η_r of $\mathbf{F}_q[X]/(g)$.*

(1) *The L-function attached to η satisfies the Euler product formula(s)*

$$L(\eta, s) = \prod_{\pi} (1 - \eta(\pi)|\pi|^{-s})^{-1}, \quad L(\eta, T) = \prod_{\pi} (1 - \eta(\pi)T^{\deg(\pi)})^{-1},$$

where π runs over monic irreducible polynomials in $\mathbf{F}_q[X]$.

(2) *If η_r is not trivial, the L-function $L(\eta, T)$ is a* polynomial *in T of degree $\leqslant \deg(g) - 1$, with constant coefficient* 1.

For the type of L-functions we consider, Part (2) was essentially first proved by F.K. Schmidt. This can be generalized considerably, though not without becoming a much harder result.

Proof. Part (1) is a simple computation which parallels the case of classical Euler products. The point is that we have

$$\eta(f_1 f_2)T^{\deg(f_1 f_2)} = \eta(f_1)T^{\deg(f_1)} \times \eta(f_2)T^{\deg(f_2)}$$

for all non-zero f_1 and f_2 so, at least formally, we can expand

$$\begin{aligned}\prod_{\pi} (1 - \eta(\pi)T^{\deg(\pi)})^{-1} &= \prod_{\pi} \sum_{d \geqslant 0} \eta(\pi^d) T^{d \deg(\pi)} \\ &= \sum_{f} \sum_{\pi_1^{d_1} \cdots \pi_m^{d_m} = f} \eta(f) T^{\deg(f)} \\ &= L(\eta, T)\end{aligned}$$

since, by unique factorization, the inner sum contains a single term for all f monic in $\mathbf{F}_q[X]$.

For Part (2), according to the definition, the coefficient of T^d in the formal power-series expansion of $L(\eta, T)$ is given by the sum

$$\sum_{\deg(f)=d} \eta(f),$$

where f runs over monic polynomials of degree d. There are, obviously, q^d such polynomials. For each, the value $\eta(f)$ depends, by definition, only on $f \pmod{g}$.

Now we have the following fact: if $d \geqslant \deg(g)$, then for every residue class $\bar{f}$ in $\mathbf{F}_q[X]/(g)$, there are exactly $q^{d-\deg(g)}$ monic polynomials of degree d in $\mathbf{F}_q[X]$ congruent to $\bar{f}$ modulo g. Indeed, if f is the unique representative of the class $\bar{f}$ which is of degree $< \deg(g)$, the required polynomials are exactly those of the type

$$f_1 = f + gh$$

where $\deg(h) = d - \deg(g)$ (they have to be of the form $f + gh$, by definition, and the condition that the degree be equal to $d \geqslant \deg(g) > \deg(f)$ implies that this degree be $d = \deg(gh) = \deg(g) + h$), and there are $q^{d-\deg(g)}$ such polynomials.

Hence we can easily compute the coefficient of T^d in $L(\eta, T)$ for all $d \geqslant \deg(g)$:

$$\begin{aligned}\sum_{\deg(f)=d} \eta(f) &= \sum_{\bar{f} \in \mathbf{F}_q[X]/(g)} \eta_r(\bar{f}) \sum_{\substack{\deg(f)=d \\ f \equiv \bar{f} \,(\mathrm{mod}\, g)}} 1 \\ &= q^{d-\deg(g)} \sum_{\bar{f} \in (\mathbf{F}_q[X]/(g))^\times} \eta_r(\bar{f}) \\ &= 0\end{aligned}$$

by orthogonality, since $\eta_r \neq 1$. It follows that $L(\eta, T)$ must be a polynomial of degree $< \deg(g)$. □

Remark 4.6. (1) There is not necessarily equality $\deg L(\eta, T) = \deg(g)$, for reasons having to do with "primitivity" (which is quite analogue to what happens for classical Dirichlet characters). For instance, for the characters associated with multiplicative character sums, since η is defined modulo $g^\flat$, the degree of the L-function is at most $\deg(g^\flat) - 1$. But, although we will see in Proposition 4.8 that in some cases there is equality, this is definitely not always the case. For example, consider $g = g_1^d g_2$ with g_1, g_2 non-constant. Then of course $\chi(Ng(x)) = \chi(Ng_2(x))$ for any $x \in \mathbf{F}_{q^\nu}$ (with $g(x) \neq 0$), so the character η can be associated to g_2 only.

(2) For completeness, lets us describe what happens if η_r is trivial (though this is not a case of interest for us). We can then compute directly the L-function by

$$\begin{aligned}L(\eta, T) &= \sum_{\substack{f \text{ monic} \\ (f,g)=1}} T^{\deg(f)} \\ &= \prod_{\pi | g} (1 - T^{\deg(\pi)}) \sum_{f \text{ monic}} T^{\deg(f)} \\ &= \prod_{\pi | g} (1 - T^{\deg(\pi)}) \sum_{\nu \geqslant 0} q^\nu T^\nu \\ &= \frac{\prod_{\pi | g} (1 - T^{\deg(\pi)})}{1 - qT}.\end{aligned}$$

Thus, although this is not a polynomial, this L-function remains a *rational function*; its only pole is located at $T = 1/q$, and in terms of the complex variable s with $T = q^{-s}$ this corresponds to poles on the line $\mathrm{Re}(s) = 1$.

Of course, a character η with η_r trivial is only primitive when $g = 1$, in which case the L-function is simply $1/(1 - qT)$.

At this point, it is natural to look more precisely at the L-function associated to a multiplicative character sum.

Proposition 4.7. *Let $\mathbf{F}_q$ be a finite field with q elements, g a non-zero monic polynomial in $\mathbf{F}_q[X]$ and χ a non-trivial multiplicative character of order $d > 1$.*

Then the L-function of the associated Dirichlet character η satisfies the exponential generating series identity

$$L(\eta, T) = \exp\Big(\sum_{\nu \geqslant 1} \frac{S_\nu(\eta)}{\nu} T^\nu\Big), \tag{4.8}$$

where

$$S_\nu(\eta) = \chi(-1)^{\nu \deg(g)} \sum_{x \in \mathbf{F}_{q^\nu}} \chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))) \tag{4.9}$$

for $\nu \geqslant 1$.

From this, we see in particular that, given g and χ, we can reconstruct the L-function of the associated character η of $\mathbf{F}_q[X]/(g)$ by means of the "companion" sums

$$\tilde{S}_\nu = \sum_{x \in \mathbf{F}_{q^\nu}} \chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x)))$$

for $\nu \geqslant 1$, for which

$$\exp\Bigl(\sum_{\nu \geqslant 1} \frac{\tilde{S}_\nu}{\nu} T^\nu\Bigr) = \exp\Bigl(\sum_{\nu \geqslant 1} \frac{S_\nu(\eta)}{\nu} (\varepsilon(\chi)T)^\nu\Bigr) = L(\eta, \varepsilon(\chi)T), \tag{4.10}$$

where $\varepsilon(\chi) = \chi(-1)^{\deg(g)}$. In doing this, we do not need to formally introduce η and its L-function. We will often use this shortcut and denote by $Z(g, \chi; T)$ the left-hand side of (4.10).

Proof. We start with the Euler product expansion of $L(\eta, T)$, computing TL'/L as a formal power series:

$$\begin{aligned} T\frac{L'}{L}(\eta, T) &= T \sum_\pi \deg(\pi) \frac{\eta(\pi) T^{\deg(\pi)-1}}{1 - \eta(\pi) T^{\deg(\pi)}} \\ &= \sum_\pi \deg(\pi) \sum_{r \geqslant 1} \eta(\pi^r) T^{r \deg(\pi)} = \sum_{\nu \geqslant 1} S_\nu T^\nu \end{aligned}$$

where

$$S_\nu = \sum_{rd = \nu} d \sum_{\deg(\pi) = d} \eta(\pi)^r.$$

Comparing with the generating series, we see that we must show that $S_\nu = S_\nu(\eta)$ as defined in (4.9). Note already that for $\nu = 1$, we have

$$S_1 = \sum_{\deg(\pi) = 1} \eta(\pi) = \sum_{t \in \mathbf{F}_q} \eta(X - t) = \chi(-1)^{\deg(g)} \sum_{x \in \mathbf{F}_q} \chi(g(x))$$

by assumption. The general case uses (4.1) in a similar way: abbreviating N instead of $N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}$, and writing π_x the minimal polynomial of an element x, we can write

$$\begin{aligned} \sum_{x \in \mathbf{F}_{q^\nu}} \chi(Ng(x)) &= \chi(-1)^{\nu \deg(g)} \sum_{x \in \mathbf{F}_{q^\nu}} \eta(\pi_x)^{\nu/\deg(x)} \\ &= \chi(-1)^{\nu \deg(g)} \sum_{\substack{\pi \text{ irred.} \\ \deg(\pi) \mid \nu}} d\eta(\pi)^{\nu/\deg(\pi)} \\ &= \chi(-1)^{\nu \deg(g)} S_\nu, \end{aligned}$$

since an irreducible (monic) polynomial π of degree dividing ν appears as the minimal polynomial of exactly d elements in $\mathbf{F}_{q^\nu}$ – its roots. □

By Proposition 4.5, for a character χ of order d such that g is not a d-th power, the L-function is a polynomial of degree $\leqslant m - 1$, $m = \deg(g^\flat)$ being the number of distinct roots of g in $\bar{\mathbf{F}}_q$. In fact, this is the exact degree under a slightly stronger condition, as follows from the following result:

Proposition 4.8. *Let $\mathbf{F}_q$ be a finite field with q elements. Let g be a non-constant monic polynomial in $\mathbf{F}_q[X]$ and let η be a Dirichlet character of $\mathbf{F}_q[X]$ modulo g, which is* primitive, *in particular has $\eta_r \neq 1$. Then the L-function $L(\eta, T)$ is a polynomial of degree* $\deg(g) - 1$, *of the form*

$$L(\eta, T) = 1 + c_1(\eta)T + \cdots + c_{d-1}(\eta)T^{d-1}$$

with leading coefficient of modulus given by

$$|c_{d-1}(\eta)| = \begin{cases} q^{(\deg(g)-1)/2}, & \text{if } \eta \mid \mathbf{F}_q^\times \text{ is non-trivial} \\ q^{\deg(g)/2-1}, & \text{otherwise,} \end{cases}$$

where the restriction of η to $\mathbf{F}_q^\times$ is defined as the restriction to non-zero constant polynomials, i.e., as the composite

$$\mathbf{F}_q^\times \to \mathbf{F}_q[X] \xrightarrow{\eta} \mathbf{C}^\times,$$

which is a multiplicative character of $\mathbf{F}_q$.

Note that although it may very well be the case that $\eta \mid \mathbf{F}_q^\times = 1$ even though η is not globally trivial (e.g., take $q = 2$; then there is non-trivial multiplicative character of $\mathbf{F}_2^\times$, but of course there are many Dirichlet characters of $\mathbf{F}_2[X]$); in that case, it should be pointed out that we have defined $\eta(0) = 0$ because the zero polynomial is not coprime with g, although the trivial character maps 0 to 1 by our convention. So one must be a little bit careful, but we will avoid evaluating $\eta \mid \mathbf{F}_q^\times$ at 0.

Proof. We will in fact give a precise "formula" for the coefficient of degree $\deg(g) - 1$ of the L-function, in terms of Gauss sums (slightly more general than those of Section 2.1), and the value of the modulus will follow from this. Below in Proposition 4.11, we will show that this is also a special case of the analogue for $L(\eta, T)$ of the functional equation (3.4) for classical Dirichlet L-functions.

For clarity, denote by R the finite ring $R = \mathbf{F}_q[X]/(g)$, and $R^\times$ its group of units. The ring R is an $\mathbf{F}_q$-vector space of dimension $d = \deg(g)$, with canonical basis given by $(1, X, \ldots, X^{d-1})$. We denote by $\ell_i(f)$, for $0 \leqslant i \leqslant d - 1$, the coefficients of $f \in R$ in this basis:

$$f = \sum_{0 \leqslant i \leqslant d-1} \ell_i(f)X^i, \qquad \text{for } f \in R.$$

Now fix a non-trivial additive character ψ of $\mathbf{F}_q$, and let ψ_1 denote the (obviously non-trivial) additive character

$$\begin{cases} R \longrightarrow \mathbf{C}^\times \\ f \mapsto \psi(\ell_{d-1}(f)). \end{cases}$$

We can now start the computation. By Definition 4.4, the coefficient $c_{d-1}(\eta)$ of T^{d-1} in $L(\eta, T)$ (which could be zero) is given by

$$c_{d-1}(\eta) = \sum_{\substack{f \text{ monic} \\ \deg(f)=d-1}} \eta(f),$$

and since the degree involved is $< \deg(g)$, so that the reduction modulo g is injective on the polynomials involved, this can be rewritten as a sum over a subset of R:

$$c_{d-1}(\eta) = \sum_{\substack{f \in R \\ \ell_{d-1}(f)=1}} \eta(f).$$

We now detect the second condition in the sum using the characters ψ and ψ_1, which are perfectly suited for this purpose:

$$\begin{aligned}c_{d-1}(\eta) = \sum_{\substack{r\in R\\ \ell_{d-1}(f)=1}} \eta(f) &= \sum_{f\in R}\eta(f)\times\frac{1}{q}\sum_{\alpha\in\mathbf{F}_q}\psi(\alpha(\ell_{d-1}(f)-1))\\ &= \frac{1}{q}\sum_{\alpha\in\mathbf{F}_q}\psi(-\alpha)\sum_{f\in R}\eta(f)\psi_1(\alpha f)\\ &= \frac{1}{q}\sum_{\alpha\in\mathbf{F}_q^\times}\psi(-\alpha)\sum_{f\in R^\times}\eta(f)\psi_1(\alpha f),\end{aligned}$$

since, on the one hand, the contribution of $\alpha = 0$ is the average of the non-trivial character η on R, and on the other-hand $\eta(f) = 0$ if $f \notin R^\times$.

The inner sum over f is quite recognizably a Gauss sum, over the finite ring R instead of a finite field. It satisfies similar properties as the Gauss sums we saw in Section 2.1, and to begin with, we have

$$\sum_{f\in R^\times}\eta(f)\psi_1(\alpha f) = \bar\eta(\alpha)\sum_{f\in R^\times}\eta(f)\psi_1(f)$$

for $\alpha \in \mathbf{F}_q^\times$, by a simple change of variable.

Thus we find the formula

$$\begin{aligned}c_{d-1}(\eta) &= \frac{1}{q}\Big(\sum_{f\in R^\times}\eta(f)\psi_1(f)\Big)\Big(\sum_{\alpha\in\mathbf{F}_q^\times}\psi(-\alpha)\bar\eta(\alpha)\Big)\\ &= \frac{\tau(\eta,\psi_1)\tau(\bar\eta\mid\mathbf{F}_q^\times,\bar\psi)}{q}.\end{aligned}\tag{4.11}$$

with the obvious notation for the Gauss sum over R.

The Gauss sum for the restriction of η to $\mathbf{F}_q^\times$ satisfies

$$|\tau(\bar\eta\mid\mathbf{F}_q^\times,\bar\psi)| = \begin{cases}1 & \text{if } \eta\mid\mathbf{F}_q^\times = 1,\\ \sqrt{q} & \text{otherwise,}\end{cases}$$

by Proposition 2.4. As for the Gauss sum over R, we claim that, for η primitive, as we assumed, we have

$$|\tau(\eta,\psi_1)|^2 = |R| = q^d.\tag{4.12}$$

Granting this, we obtain

$$|c_{d-1}(\eta)| = \begin{cases}q^{(d-1)/2} & \text{if } \eta\mid\mathbf{F}_q^\times \neq 1,\\ q^{d/2-1} & \text{otherwise,}\end{cases}$$

as desired.

To prove (4.12), let us assume first that g is squarefree, in which case the primitivity assumption is more transparent. The easiest way to proceed is to use once more the isomorphism

$$R \simeq \prod_{\pi\mid g}\mathbf{F}_q[X]/(\pi),$$

from which it follows that η factors as

$$\eta(f)=\prod_{\pi\mid g}\eta_\pi(f),$$

where η_π is a character of the multiplicative group $(\mathbf{F}_q[X]/\pi)^\times$ (of a finite field with $q^{\deg(\pi)}$ elements); then primitivity means that *none of the* η_π is trivial. Using the above isomorphism again, we derive by the Chinese Remainder Theorem the product formula

$$\tau(\eta,\psi_1)=\prod_{\pi\mid g}\tau(\eta_\pi,\psi_\pi),$$

where ψ_π is the π-component of the additive character ψ_1 of R. Once we show that ψ_π is also non-trivial for all π, this product formula leads to the claim after applying Proposition 2.4 once more. But indeed, for $x\in\mathbf{F}_{q^{\deg(\pi)}}$, we compute $\psi_\pi(x)$ as $\psi_1(f)$ where f is a polynomial in $\mathbf{F}_q[X]$ mapping a root (denoted α_π) of π to x and mapping the roots of the other irreducible factors to 0. Observe that any of the monic polynomials f given by

$$f=\frac{g}{\pi}h$$

where $\deg(h)=\deg(\pi)-1\geqslant 0$ satisfy the latter condition. We then have $\psi_1(f)=1$, and more generally for such a polynomial and any $x\in\mathbf{F}_q^\times$, we get

$$\psi_\pi(xf(\alpha_\pi))=\psi_1(xf)=\psi(x).$$

Now, since ψ is non-trivial, we can find α for which this is not $=1$. □

Corollary 4.9. *Let $\mathbf{F}_q$ be a finite field with q elements, let g be a non-constant monic polynomial and χ a character of order d such that $d\nmid\deg(g)$ and there is no irreducible factor π of g of* multiplicity *divisible by d.*

(1) *The zeta function $Z(g,\chi;T)$ is a polynomial of degree $m-1$, taking value 1 at $T=0$, and with leading coefficient of modulus $q^{(m-1)/2}$, where m is the number of distinct zeros of g in $\bar{\mathbf{F}}_q$.*

(2) *Factor the polynomial as*

$$Z(g,\chi;T)=\prod_{1\leqslant j\leqslant m-1}(1-\Omega_jT),$$

with $\Omega_j\in\mathbf{C}$. Then for any $\nu\geqslant 1$, we have

$$\tilde{S}_\nu=\sum_{x\in\mathbf{F}_{q^\nu}}\chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x)))=-\Big\{\Omega_1^\nu+\cdots+\Omega_{m-1}^\nu\Big\},\tag{4.13}$$

and moreover $|\Omega_1\cdots\Omega_{m-1}|=q^{(m-1)/2}$.

Proof. For (1), we apply Proposition 4.8 to the character η associated to the data (g,χ). The assumption that $\deg(g)\not\equiv 0\,(\mathrm{mod}\,d)$ implies that $\chi^{\deg(g)}\neq 1$, so that g is not a d-th power in $\bar{\mathbf{F}}_q[X]$. Also, η restricted to the non-zero constants is the same as $\chi^{\deg(g)}$ by (4.7), hence is non-trivial.

We know that η is defined modulo $g^\flat$ (defined in Corollary 4.3), which has degree m. To check that it is indeed *primitive* modulo $g^\flat$, we argue as in the proof of this corollary that, for π dividing g, the π-component η_π is the character of $(\mathbf{F}_q[X]/\pi)^\times$ given by

$$\eta_\pi(x)=\chi(N_{\mathbf{F}_{q^{\deg(\pi)}}/\mathbf{F}_q}f(\alpha_\pi)^{k_\pi})$$

for $x\in(\mathbf{F}_q[X]/\pi)^\times$ and f a polynomial in $\mathbf{F}_q[X]$ mapping the root α_π of π to x and roots of the other irreducible factors to 1. Thus, the proof in Corollary 4.3 exactly shows that one can find x with $\eta_\pi(x) \neq 1$ for all those π where the multiplicity k_π is not divisible by d, which we assumed were all irreducible factors.

(2) We now consider the identity

$$\exp\Bigl(\sum_{\nu\geqslant 1}\frac{\tilde{S}_\nu}{\nu}T^\nu\Bigr) = Z(g,\chi;T) = \prod_{1\leqslant j\leqslant m-1}(1-\Omega_j T),$$

and proceed to apply the operator TZ'/Z on both sides; applying the geometric series expansion, we obtain

$$\begin{aligned}\sum_{\nu\geqslant 1}\tilde{S}_\nu T^\nu &= T\sum_{1\leqslant j\leqslant m-1}(-\Omega_j)\sum_{\nu\geqslant 0}\Omega_j^\nu T^\nu\\ &= -\sum_{\nu\geqslant 1}\Bigl(\sum_{1\leqslant j\leqslant m-1}\Omega_j^\nu\Bigr)T^\nu,\end{aligned}$$

hence the result after comparing.

As to the product of the Ω_j's, up to sign, this is the same as the leading term of the zeta function, hence its modulus is given by Proposition 4.8. □

Remark 4.10. (1) We have already noticed, taking examples like $g = g_1^d g_2$ with g_i non-constant, that the restrictions in this corollary are not artificial.

(2) In terms of the L-function $L(\eta,T)$, if we factor the latter as

$$L(\eta,T) = \prod_{1\leqslant j\leqslant m-1}(1-\omega_j T),$$

as we can since $L(\eta,0)=1$, we get

$$\sum_{x\in\mathbf{F}_{q^\nu}}\chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))) = -\varepsilon(\chi)^\nu\Bigl\{\omega_1^\nu+\cdots+\omega_{m-1}^\nu\Bigr\}.$$

(3) If g does not satisfy the assumption of this corollary, but still is associated with a non-trivial η, we can still use Proposition 4.5 to deduce a representation

$$\sum_{x\in\mathbf{F}_{q^\nu}}\chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))) = -\varepsilon(\chi)^\nu\Bigl\{\omega_1^\nu+\cdots+\omega_\delta^\nu\Bigr\}, \tag{4.14}$$

for all ν, where $\delta\leqslant m-1$ is the degree of the L-function $L(\eta,T)$ as polynomial, and the $\omega_j\neq 0$ are such that

$$L(\eta,T) = (1-\omega_1 T)\cdots(1-\omega_\delta T).$$

It will be often convenient to define $\omega_j = 0$ for $\delta+1\leqslant j\leqslant m-1$, so that the formula becomes the same as the one before, but some of the inverse roots ω_j can be zero.

(4) Examples of situations where this corollary applies are given by χ of order 2 and g squarefree of odd degree.

For the purpose of proving Theorems 3.1, we do not need more than Proposition 4.5 (even the knowledge of the leading term is only useful to give additional information). However, it is of some interest to go deeper, especially by analogy with classical Dirichlet characters and their functional equation (3.4). This has, indeed, an analogue, which we prove in the simplest case:

Proposition 4.11. *Let $\mathbf{F}_q$ be a finite field with q elements. Let g be a non-constant monic polynomial of degree d in $\mathbf{F}_q[X]$ and let η be a Dirichlet character of $\mathbf{F}_q[X]$ modulo*

g, which is primitive *and has $\eta \mid \mathbf{F}_q^\times \neq 1$. Then the L-function $L(\eta, T)$ satisfies the functional equation*

$$L(\eta, T) = W(\eta) T^{d-1} L(\bar{\eta}, (qT)^{-1}),$$

or equivalently

$$L(\eta, q^{-s}) = W(\eta) Q^{-s} L(\bar{\eta}, q^{-(1-s)}), \qquad Q = q^{d-1},$$

where

$$W(\eta) = \frac{\tau(\eta, \psi_1)\tau(\bar{\eta} \mid \mathbf{F}_q^\times, \bar{\psi})}{q}.$$

Proof. The argument is an elaboration of that of Proposition 4.8, which follows indeed from the functional equation by comparing the leading terms on both sides. So we use notation like in the proof above, in particular we consider $R = \mathbf{F}_q[X]/(g)$. Since $\bar{\eta}$ is itself clearly a primitive Dirichlet character with $\bar{\eta}|\mathbf{F}_q^\times \neq 1$, we can write the functional equation as a polynomial identity

$$1 + c_1(\eta)T + \cdots + c_{d-1}(\eta)T^{d-1} = W(\eta)T^{d-1}\Big(1 + \frac{c_1(\bar{\eta})}{qT} + \cdots + \frac{c_{d-1}(\bar{\eta})}{(qT)^{d-1}}\Big), \tag{4.15}$$

where

$$c_j(\eta) = \sum_{\substack{f \text{ monic} \\ \deg(f)=j}} \eta(f),$$

and similarly for $\bar{\eta}$. In particular, Proposition 4.8 shows that the coefficients of T^{d-1} on both sides coincide.

Fix j with $1 \leqslant j \leqslant d-1$. The reduction map $\mathbf{F}_q[X] \to R$ is injective on polynomials of degree $\leqslant j \leqslant d-1$. We express $c_j(\eta)$ as a sum over R, and then $R^\times$, by detecting with additive characters the $d-j$ conditions

$$\ell_j(f) = 1, \quad \ell_{j+1}(f) = \cdots = \ell_{d-1}(f) = 0,$$

which characterize the image of monic polynomials of degree j in R. Doing this, we find that $c_{j-1}(\eta)$ is equal to

$$\frac{1}{q^{d-j}} \sum_{f \in R} \eta(f) \sum_{\boldsymbol{\alpha} = (\alpha_j, \ldots, \alpha_{d-1}) \in \mathbf{F}_q^{d-j}} \psi(\alpha_j(\ell_j(f) - 1) + \alpha_{j+1}\ell_{j+1}(f) + \cdots)$$

which we write as

$$\frac{1}{q^{d-j}} \sum_{\boldsymbol{\alpha}} \psi(-\alpha_j) S(\eta, \boldsymbol{\alpha}),$$

with

$$S(\eta, \boldsymbol{\alpha}) = \sum_{f \in R} \eta(f) \psi\Big(\sum_{k=j}^{d-1} \alpha_k \ell_k(f)\Big) = \sum_{f \in R^\times} \eta(f) \psi\Big(\sum_{k=j}^{d-1} \alpha_k \ell_k(f)\Big).$$

Now we observe that, in R, we have

$$\sum_{k=j}^{d-1} \alpha_k \ell_k(f) = \ell_{d-1}(f h_{\boldsymbol{\alpha}})$$

for the polynomial (where the coefficients of $\boldsymbol{\alpha}$ are coded in *descending* order) given by

$$h_{\boldsymbol{\alpha}} = \alpha_j X^{d-1-j} + \cdots + \alpha_{d-2} X + \alpha_{d-1}.$$

We can therefore interpret $S(\eta,\boldsymbol{\alpha})$ as another Gauss sum over R using the characters $\psi_h(f)=\psi(\ell_{d-1}(fh))=\psi_1(fh)$, namely

$$S(\eta,\boldsymbol{\alpha})=\tau(\eta,\psi_{h_{\boldsymbol{\alpha}}}).$$

We now claim that, as in (2.2), we have

$$\tau(\eta,\psi_h)=\bar{\eta}(h)\tau(\eta,\psi_1) \tag{4.16}$$

for any $h\in R$ (not necessarily in $R^\times$). Taking this for granted, we get

$$\begin{aligned} c_j(\eta) &= \frac{\tau(\eta,\psi_1)}{q^{d-j}}\sum_{\boldsymbol{\alpha}}\overline{\eta(h_{\boldsymbol{\alpha}})}\psi(-\alpha_j) \\ &= \frac{\tau(\eta,\psi_1)}{q^{d-j}}\sum_{\substack{h\in\mathbf{F}_q[X]\\ \deg(h)\leqslant d-1-j}}\overline{\eta(h)}\psi_{X^j}(-h) \end{aligned}$$

after identifying polynomials of degree $\leqslant d-1-j\leqslant d-1$ with the polynomials $h_{\boldsymbol{\alpha}}$ via their coefficients.

For the final step, we write

$$\sum_{\substack{g\in\mathbf{F}_q[X]\\ \deg(h)\leqslant d-1-j}}\overline{\eta(h)}\psi_{X^j}(-h)=\sum_{\substack{h\in\mathbf{F}_q[X]\\ \deg(h)\leqslant d-2-j}}\overline{\eta(h)}+\sum_{\deg(h)=d-1-j}\overline{\eta(h)}\psi_{X^j}(-h),$$

and we notice first that

$$\begin{aligned} \sum_{\deg(h)=d-1-j}\overline{\eta(h)}\psi_{X^j}(-h) &= \sum_{\alpha\in\mathbf{F}_q^\times}\psi(-\alpha)\sum_{\substack{\deg(h)=d-1-j\\ \ell_{d-1-j}(h)=\alpha}}\overline{\eta(h)} \\ &= \sum_{\alpha\in\mathbf{F}_q^\times}\psi(-\alpha)\bar{\eta}(\alpha)\sum_{\substack{\deg(h)=d-1-j\\ h\text{ monic}}}\overline{\eta(h)} \end{aligned}$$

(by a substitution $h\mapsto\alpha^{-1}h$)

$$=\tau(\bar{\eta}\mid\mathbf{F}_q^\times,\bar{\psi})c_{d-1-j}(\bar{\eta}).$$

The remaining term is

$$R=\sum_{\substack{h\in\mathbf{F}_q[X]\\ \deg(h)\leqslant d-2-j}}\overline{\eta(h)},$$

and it follows by substituting h by αh with $\alpha\in\mathbf{F}_q^\times$ that

$$R=\overline{\eta(\alpha)}R$$

for every such α. Since we assumed that $\bar{\eta}\mid\mathbf{F}_q^\times\neq 1$, this implies that $R=0$, and the final relations

$$c_j(\eta)=\frac{\tau(\eta,\psi_1)\tau(\bar{\eta}\mid\mathbf{F}_q^\times,\psi)}{q^{d-j}}c_{d-1-j}(\bar{\eta}),$$

valid for $1\leqslant j\leqslant d-1$, are precisely those that are needed to check the functional equation (4.15) by comparison of coefficients.

We must still check (4.16). If $h\in R^\times$ (i.e., if h is coprime with g in $\mathbf{F}_q[X]$), this is obvious by the usual change of variable. If $h\notin R^\times$, the right-hand side is zero, and we must show that $\tau(\eta, \psi_h)$ is also zero. But assuming first that g is squarefree, we can factor the Gauss sum as

$$\tau(\eta, \psi_h) = \prod_{\pi \mid g} \tau(\eta_\pi, \psi_{\pi,h})$$

with $\psi_{\pi,h}(x) = \psi_\pi(h(\alpha_\pi)x)$ for $x \in \mathbf{F}_q(\alpha_\pi)$ (with notation as in the proof of Proposition 4.8). If h is not coprime with g, one of the irreducible factors π divides h, and then $\psi_{\pi,h}$ is the trivial additive character. Since $\tau(\eta_\pi, 1) = 0$ (because η_π is non-trivial by primitivity), we obtain the desired conclusion. □

Example 4.12. In some cases, the leading term (i.e., the constant $W(\eta)$) of Proposition 4.8 can be computed more explicitly. We give one important example, corresponding to the case of Corollary 3.3. Let q be odd, let $g \in \mathbf{F}_q[X]$ be a cubic polynomial of the form

$$g = X^3 + aX + b$$

which is squarefree in $\mathbf{F}_q[X]$, and let χ be the non-trivial quadratic character of $\mathbf{F}_q$. Since $m = d = 3$, the L-function for the corresponding Dirichlet character is of the form

$$L(\eta, T) = 1 + c_1(\eta)T + c_2(\eta)T^2,$$

for some $c_1(\eta) \in \mathbf{Z}$ (indeed, we have

$$c_1(\eta) = \sum_{t \in \mathbf{F}_q} \eta(X - t) = \chi(-1) \sum_{t \in \mathbf{F}_q} \chi(g(t))$$

and the values of χ are integers), and a leading term with $|c_2(\eta)| = q$. Moreover, this leading term is an integer, since we have again the expression

$$c_2(\eta) = \chi(-1) \sum_{a_0, a_1 \in \mathbf{F}_q} \eta(X^2 + a_1 X + a_0),$$

where, denoting α_1, α_2, α_3 the three roots of g in $\bar{\mathbf{F}}_q$, we have

$$\eta(X^2 + a_1 X + a_0) = \chi\Big(\prod_{1 \leqslant i \leqslant 3} (\alpha_i^2 + a_1 \alpha_i + a_0) \Big)$$

by (4.5). Therefore, $c_2(\eta)$ is either q or $-q$, and we have the question of determining the sign (a variant of the question of computing the sign of a quadratic Gauss sum).

4.2. The Riemann hypothesis for multiplicative character sums

We consider the data (g, χ) over a finite field $\mathbf{F}_q$, for which we have the associated multiplicative character sum. Comparing (4.14) with the statement of Theorem 3.1, we see that the latter would follow immediately if we knew suitable bounds for the moduli of the Ω_j's, or equivalently of the ω_j's. This is indeed the “right” version of the Riemann Hypothesis.

Theorem 4.13 (A. Weil). *Let $\mathbf{F}_q$ be a finite field, g a non-constant monic polynomial in $\mathbf{F}_q[X]$.*

(1) *If η is a primitive Dirichlet character modulo g, non-trivial on the subgroup of constants, then writing*

$$L(\eta, T) = \prod_{1 \leqslant j \leqslant \deg(g) - 1} (1 - \omega_j T),$$

every ω_j is a q-Weil number of weight 1. In particular, every zero ρ of $L(\eta, q^{-s})$ satisfies $\mathrm{Re}(\rho) = 1/2$.

(2) *More generally, for any non-trivial η, if its L-function is a polynomial of degree $\delta \leqslant \deg(g)-1$ written as*

$$L(\eta,T)=\prod_{1\leqslant j\leqslant\delta}(1-\omega_jT),$$

we have

$$|\omega_j|\leqslant\sqrt{q}$$

for all j.

We will not prove this in full generality here but restrict our attention for the moment to those η coming from multiplicative exponential sums.

Proposition 4.14 (A. Weil). *Let $\mathbf{F}_q$ be a finite field, g a non-constant monic polynomial in $\mathbf{F}_q[X]$, χ a multiplicative character of order $d>1$.*

(1) *If $d\nmid\deg(g)$ and no irreducible factor of g is of multiplicity divisible by d, then*

$$Z(g,\chi;T)=\prod_{1\leqslant j\leqslant\deg(g^\flat)-1}(1-\Omega_jT),$$

every Ω_j is a q-Weil number of weight 1.

(2) *More generally, if g is not a d-th power in $\bar{\mathbf{F}}_q[X]$, then if its zeta function is a polynomial of degree $\delta\leqslant\deg(g^\flat)-1$ written as*

$$\prod_{1\leqslant j\leqslant\delta}(1-\Omega_jT),$$

we have

$$|\Omega_j|\leqslant\sqrt{q}\tag{4.17}$$

for all j.

Using the formula (4.14) for the exponential sums in terms of the Ω_j's, we see that this proposition immediately implies Theorem 3.1.

Before proving this, we first observe that it is enough to prove Part (2); indeed, in the situation of Part (1), we know that $\delta=\deg(g^\flat)-1$ and that

$$|\Omega_1\cdots\Omega_\delta|=q^{(\delta-1)/2}$$

by Proposition 4.8, so that, for any j, we have

$$|\Omega_j|\geqslant\frac{q^{(\delta-1)/2}}{\prod_{k\neq j}|\Omega_k|}\geqslant q^{1/2}$$

by (4.17). Hence, in fact, we have $|\Omega_j|=\sqrt{q}$ exactly. There remains to check that Ω_j is a Weil number, i.e., that all its conjugates are also of modulus $\sqrt{q}$. More generally, if η is a primitive Dirichlet character modulo g, non-trivial on the constants, and σ is an automorphism of $\mathbf{C}$ (or of the field generated by the roots of the L-function of η), then we have

$$L(\eta,T)^\sigma=\sum_{f\text{ monic}}(\sigma\circ\eta)(f)T^{\deg(f)}=L(\sigma\circ\eta,T),$$

and since $\sigma\circ\eta$ is again easily seen to be a primitive character modulo g, non-trivial on the constants, we see the conjugates under σ of the inverse roots ω_j of $L(\eta,T)$ are inverse roots of $L(\sigma\circ\eta,T)$. Hence in our case, applying the argument above to $\sigma\circ\eta$ ensures that the image under σ of Ω_j is of modulus $\sqrt{q}$, for any σ.

Now, we will use the following nice trick (with $B=\sqrt{q}$) in order to prove (4.17) without having to deal with the individual ω_j's:

Lemma 4.15. *Let $\delta \geqslant 1$ be an integer, and $\omega_1, \ldots, \omega_\delta$ complex numbers such that, for some constant $A \geqslant 0$, we have*

$$|\omega_1^\nu + \cdots + \omega_\delta^\nu| \leqslant AB^\nu$$

for all $\nu \geqslant 1$. Then in fact

$$|\omega_j| \leqslant B$$

for all j.

Proof. Although this is not necessary, and a bit wasteful, we use the "slick" proof: first, we can assume all $\omega_j \neq 0$; we then consider the generating series

$$f(z) = \sum_{\nu \geqslant 1} (\omega_1^\nu + \cdots + \omega_\delta^\nu) z^\nu$$

as a function of $z \in \mathbf{C}$. The assumption shows that f is an absolutely convergent power series, hence a holomorphic function, in the open disc $|z| < B^{-1}$. On the other hand, we can write

$$f(z) = \sum_{j=1}^{\delta} \sum_{\nu \geqslant 1} \omega_j^\nu z^\nu = \sum_{j=1}^{\delta} \frac{\omega_j z}{1 - \omega_j z}$$

at least in the open disc

$$|z| < \min_j |\omega_j|^{-1}.$$

By the principle of analytic continuation, the rational function on the right-hand side must have radius of convergence equal to that of f, i.e., at least $1/B$. Hence $|\omega_j|^{-1} \geqslant B^{-1}$, which gives the result. □

The next step before embarking on the difficult part of the work, is to exploit this lemma and averaging over characters to reduce to some point-counting. This will be crucial, as it allows us to use positivity arguments.

Lemma 4.16 (Reduction to point counting). *Let $\mathbf{F}_q$ be a finite field, g a non-constant monic polynomial in $\mathbf{F}_q[X]$, $d \mid q-1$ an integer. For all $\nu \geqslant 1$, we have*

$$\sum_{\substack{\chi^d = 1 \\ \chi \neq 1}} \sum_{x \in \mathbf{F}_{q^\nu}} \chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q} g(x)) = |\{(x, y) \in \mathbf{F}_{q^\nu} \times \mathbf{F}_{q^\nu} \mid y^d = g(x)\}| - q^\nu,$$

where χ runs over non-trivial multiplicative characters of $\mathbf{F}_q$ of order dividing d.

Proof. This is immediate by the formula (1.7), after subtracting the contribution of the trivial character, once we note that for any $\nu \geqslant 1$, the multiplicative characters

$$\chi \circ N : x \mapsto \chi(N_{\mathbf{F}_{q^\nu}/\mathbf{F}_q} x)$$

of $\mathbf{F}_{q^\nu}$ are all the characters of order dividing d (which divides $q^\nu - 1$) of the extension field. It is clear that

$$\chi \mapsto \chi \circ N$$

is a homomorphism from the group of characters of order dividing d of $\mathbf{F}_q$ to that for $\mathbf{F}_{q^\nu}$. Since each of these groups is cyclic of order d, it is enough to show that if $\chi \neq 1$, then $\chi \circ N \neq 1$. However, this follows immediately from the surjectivity of the norm map (Lemma 1.3). □

The main theorem will then be the following:

Theorem 4.17 (Stepanov, Bombieri). *Let $\mathbf{F}_q$ be a finite field* where $q=p^{2\nu}$ for some $\nu \geqslant 1$, *$g \in \mathbf{F}_q[X]$ a non-constant polynomial, $d \mid q-1$ an integer such that* $(\deg(g), d)=1$. *Then there exists a constant $C \geqslant 0$, depending only on d and the degree of g, such that*

$$||\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^d = g(x)\}|-q| \leqslant Cq^{1/2}.$$

Here is how all these ingredients combine, in the case where g has degree coprime with d and is squarefree: applying (4.14) to all the data (g,χ) where χ is non-trivial and of order dividing d, we find a finite collection of complex numbers containing all ω_j's for such a χ, with the property that

$$\Bigl|\sum_i \omega_i^{2\nu}\Bigr| \leqslant Cq^{\nu}$$

for all ν (by Theorem 4.17 applied to g over $\mathbf{F}_{q^{2\nu}}$, using the fact that C is independent of q). By Lemma 4.15, each of the ω_i's satisfy $|\omega_i|^2 \leqslant q$, hence $|\omega_i| \leqslant \sqrt{q}$, as we wanted to prove.

4.3. Stepanov's method

We now finally attack the proof of Theorem 4.17, using notation as in the statement. The original basic idea of Stepanov is so simple that it is quite surprising that it should work: he constructed an auxiliary polynomial $A \in \mathbf{F}_q[X]$, and a parameter $m \geqslant 1$, such that (i) $A \neq 0$; (ii) for any x such that $g(x)$ is a d-th power in $\mathbf{F}_q$, A vanishes at x *to order at least m*, i.e., the polynomial $(X-x)^m$ divides A. Since the number of solutions to any equation $y^d=\beta$, $\beta \in \mathbf{F}_q$, is at most d, it follows by looking at the degree of A that

$$|\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^d = g(x)\}| \leqslant \frac{d\deg(A)}{m}. \tag{4.18}$$

This seems to be only half (and maybe the easier half!) of the work, as it provides only an upper bound for the point-counting. However, a well-known trick shows this is enough. We phrase it as a lemma:

Lemma 4.18 (From upper bound to lower bound). *Let $\mathbf{F}_q$ be a finite field. For any $d \mid q-1$ and $g \in \mathbf{F}_q[X]$ non-constant of degree coprime with d, let $a(g) \in \mathbf{Z}$ be defined by the equality*

$$|\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^d = g(x)\}| = q + a(g).$$

Then we have

$$|\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^d = g(x)\}| \geqslant q-(d-1)\max_{\varepsilon \in \mathbf{F}_q^\times} |a(\varepsilon g)|.$$

Proof. The idea is simply to cover all of $\mathbf{F}_q$ with d sets corresponding to the x-coordinates of solution sets for various polynomials of the same degree as g. This is possible because, for every $x \in \mathbf{F}_q$, $f(x)$ is, if not a d-th power, at least one up to multiplication by a coset representative of the finite cyclic group

$$\mathbf{F}_q^\times/(\mathbf{F}_q^\times)^d.$$

To be precise, fix representatives $\{1,\varepsilon_2,\ldots,\varepsilon_d\}$ of this group, and for any of them ε, let

$$h_\varepsilon = \varepsilon^{-1}g,$$

which has same degree as g, and let

$$\mathcal{C}_\varepsilon^* = \{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q^\times \mid y^d = h_\varepsilon(x)\}\ ;$$

observe that y is taken to be non-zero here so that, by definition, the cardinality of this set can be written

$$|\mathcal{C}_\varepsilon^*| = q + a(h_\varepsilon) - z(g),$$

where $0 \leqslant z(g) \leqslant \deg(g)$ is the number of zeros of g in $\mathbf{F}_q$.

Then, since $g(x)$ is of the form εy^d for a unique coset representative ε and for d values of y, provided $g(x) \neq 0$, we have

$$\sum_\varepsilon |\mathcal{C}_\varepsilon^*| = \sum_\varepsilon \sum_{y^d=\varepsilon^{-1}g} 1 = \sum_{\substack{x\in\mathbf{F}_q\\ g(x)\neq 0}} d = d(q - z(g)).$$

Comparing, this means that

$$d(q-z(g)) + \sum_\varepsilon a(h_\varepsilon) = d(q-z(g)),$$

hence

$$a(g) \geqslant -(d-1)\max_{\varepsilon\neq 1} |a(g_\varepsilon)|,$$

which is the claim of the lemma. □

This lemma shows that it will be enough to prove the upper bound

$$|\{(x,y)\in\mathbf{F}_q\times\mathbf{F}_q \mid y^d = g(x)\}| \leqslant q + Cq^{1/2}. \tag{4.19}$$

(for q a square and some constant C depending only on d and $\deg(g)$) in order to finally prove Theorem 4.17: applied to an arbitrary $h = \varepsilon g$ with $\varepsilon \in \mathbf{F}_q^\times$, we obtain $a(h) \leqslant Cq^{1/2}$, and hence

$$|\{(x,y)\in\mathbf{F}_q\times\mathbf{F}_q \mid y^d = g(x)\}| \geqslant q - (d-1)Cq^{1/2},$$

which together with (4.19) gives Theorem 4.17.

However, implementing this strategy leads an annoying technical issue: how should one detect zeros of high order of polynomials in positive characteristic p? Over $\mathbf{Q}$, one would of course say that $P \in \mathbf{Z}[X]$ is divisibly by $(X-x)^m$ if and only if

$$P(x) = P'(x) = \cdots = P^{(m-1)}(x) = 0.$$

However, this fails for zeros of order $m \geqslant p$ in characteristic p: indeed, the polynomial $P = X^p \in \mathbf{F}_p[X]$ is such that $P^{(j)}(0) = 0$ for *all* $j \geqslant 0$ – because the p-th derivative is equal to the constant $p! = 0$ in $\mathbf{F}_p$ (though not in $\mathbf{Q}$). Since, as it turns out, the parameter m in the proof of Theorem 4.17 will often be larger than the characteristic p, Stepanov required another tool, the *Hasse derivatives*.

E. Bombieri, in his Bourbaki report on Stepanov's work [1], gave a proof of a generalized version of Stepanov's result, using a simpler construction avoiding the use of Hasse derivatives. We give this argument, specialized to the situation at hand. This has the further advantage of bringing in, quite naturally and elementarily, some notions of algebraic geometry.

Bombieri's idea depends on considering the set of solutions of the equation as a geometric object (an algebraic curve); thus, let

$$\mathcal{C} = \{(x,y)\in\bar{\mathbf{F}}_q^2 \mid y^d = g(x)\} \tag{4.20}$$

be this set of solutions, with coordinates in a fixed algebraic closure of $\mathbf{F}_q$. This is an infinite set, and one denotes

$$\mathcal{C}(\mathbf{F}_q) = \{(x,y)\in\mathbf{F}_q^2 \mid y^d = g(x)\} = \{\boldsymbol{x}\in\mathcal{C} \mid \mathrm{Fr}(\boldsymbol{x}) = \boldsymbol{x}\},$$

where the Frobenius automorphism Fr acts on $\mathcal{C}$ in the obvious way, namely:

$$\mathrm{Fr}(x,y)=(x^q,y^q)$$

Note that, to obtain the last crucial identity

$$\mathcal{C}(\mathbf{F}_q)=\{\boldsymbol{x}\in\mathcal{C}\mid \mathrm{Fr}(\boldsymbol{x})=\boldsymbol{x}\},$$

we used the fact that $g\in\mathbf{F}_q[X]$, so that $\mathrm{Fr}(g(x))=g(\mathrm{Fr}(x))$.

Instead of polynomials, we consider functions on $\mathcal{C}$, and more precisely, restrictions to $\mathcal{C}$ of polynomials in $\mathbf{F}_q[X,Y]$. The ring of such functions can be identified with the quotient ring

$$\mathcal{O}(\mathcal{C})=\mathbf{F}_q[X,Y]/(Y^d-g(X)),$$

(which is easily seen to be an integral domain). Any element $f\in\mathcal{O}(\mathcal{C})$ can be expressed uniquely as

$$f=\sum_{i=0}^{d-1}g_i(X)Y^i, \tag{4.21}$$

where $g_i\in\mathbf{F}_q[X]$. The auxiliary function will be constructed as an element of $\mathcal{O}(\mathcal{C})$. A suitable analogue of (4.18) is obtained by looking at functions $0\neq f\in\mathcal{O}(\mathcal{C})$ which vanish to some larger order $\geqslant m$ at all points $\boldsymbol{x}\in\mathcal{C}(\mathbf{F}_q)$, and yet have a small enough "degree". One of the main points is indeed to define this degree for these functions which are more complicated than mere polynomials. A suitable definition is as follows, at least in the case $(d,\deg(g))=1$, which we now assume:

Definition 4.19 (Degree "at infinity" of a function on $\mathcal{C}$). Let $\mathbf{F}_q$, $d\mid q-1$ and $g\in\mathbf{F}_q[X]$ be as above, with $(d,\deg(g))=1$. Then for $0\neq f\in\mathcal{O}(\mathcal{C})$, expressed as (4.21), we define

$$\deg(f)=\max\{d\deg(g_i)+i\deg(g)\mid 0\leqslant i\leqslant d-1\}\geqslant 0,$$

with the convention that $\deg(0)=-\infty$, so the index ranges implicitly only over those i such that $g_i\neq 0$.

The geometric idea is the following: on the curve $\mathcal{C}$, if the point $\boldsymbol{x}\in\mathcal{C}$ "goes to infinity", the function f has a pole of order $\deg(f)$ at infinity. For instance, we have

$$\deg(X)=d,\qquad \deg(Y)=\deg(g).$$

Here are the basic properties of this degree, which confirm the intuitive ideas above:

Proposition 4.20 (Degree and zeros of functions). *Let $\mathbf{F}_q$ be a finite field with q elements, $d\mid q-1$ with $d>1$ an integer, $g\in\mathbf{F}_q[X]$ a non-constant polynomial with* $(\deg(g),d)=1$ and g squarefree. *Let $\mathcal{C}$ be the curve* (4.20).

(1) *For all f_1, f_2, both non-zero, in $\mathcal{O}(\mathcal{C})$, we have*

$$\deg(f_1f_2)=\deg(f_1)+\deg(f_2), \tag{4.22}$$

and if $f_1+f_2\neq 0$, then

$$\deg(f_1+f_2)\leqslant\max(\deg(f_1),\deg(f_2)). \tag{4.23}$$

(2) *For any $0\neq f\in\mathcal{O}(\mathcal{C})$, let $\deg_0(f)$ denote the number of zeros of f on $\mathcal{C}$,* counted with multiplicity*; then we have*

$$\deg_0(f)\leqslant\deg(f).$$

In particular, if all zeros of f, with at most m_0 exceptions, have multiplicity $\geqslant m$, we have

$$|\{\boldsymbol{x} \in \mathcal{C} \mid f(\boldsymbol{x}) = 0\}| \leqslant \frac{\deg(f)}{m} + m_0. \tag{4.24}$$

For the second part, we must provide the definition of zeros of f and their multiplicity. Of course, the former is clear: $\boldsymbol{x} \in \mathcal{C}$ is a zero of f if $f(\boldsymbol{x}) = 0$. The multiplicity can be defined in great generality, but we take the following convenient concrete description.

Definition 4.21 (Multiplicity of a zero). Let $\boldsymbol{x} = (x, y) \in \mathcal{C}$ be any point.

(1) If $y \neq 0$, $\boldsymbol{x}$ is *a zero of f with multiplicity $\geqslant m$* if and only if the element

$$\frac{f}{(X-x)^m}$$

of the fraction field of $\mathcal{O}(\mathcal{C})$ is defined at $\boldsymbol{x}$, i.e., can be written f_1/f_2 with f_1, $f_2 \in \mathcal{O}(\mathcal{C})$ and $f_2(x, y) \neq 0$.

(2) If $y = 0$, $\boldsymbol{x} = (x, 0)$ is a zero of f with multiplicity $\geqslant m$ if and only if

$$\frac{f}{Y^m}$$

is defined at $\boldsymbol{x}$, in the same sense.

Note that clearly $f(\boldsymbol{x}) = 0$ if $\boldsymbol{x}$ is a zero of multiplicity $\geqslant 1$. It is important for us to know that the converse holds, and this is not entirely obvious.

Lemma 4.22. *With notation and assumption as above, if $\boldsymbol{x} = (x, y) \in \mathcal{C}$ is a zero of $f \in \mathcal{O}(\mathcal{C})$, then the multiplicity of f at $\boldsymbol{x}$ is $\geqslant 1$.*

Proof. Assume first that $y \neq 0$. Then we write

$$g_i = (X - x)h_i + g_i(x), \qquad 0 \leqslant i \leqslant d-1, \quad h_i \in \mathbf{F}_q[X],$$

(which is Euclidean division), to get

$$\frac{f}{X-x} = \sum_{i=0}^{d-1} h_i Y^i + \frac{\sum_{i=0}^{d-1} g_i(x) Y^i}{X-x},$$

and then further we write

$$Y^i = y^i + (Y - y)\tilde{h}_i(Y), \qquad \tilde{h}_i \in \mathbf{F}_q[Y],$$

to obtain

$$\begin{aligned}\frac{f}{X-x} &= \sum_{i=0}^{d-1} h_i Y^i + \frac{f(x,y)}{X-x} + \Big(\sum_{i=0}^{d-1} \tilde{h}_i(Y)\Big)\frac{Y-y}{X-x} \\ &= \sum_{i=0}^{d-1} h_i Y^i + \Big(\sum_{i=0}^{d-1} \tilde{h}_i(Y)\Big)\frac{Y-y}{X-x}\end{aligned}$$

since $f(x, y) = 0$ by assumption. Finally, using the fact that $y^d = g(x)$, we use the relation

$$Y^d - y^d = g(X) - g(x)$$

to derive

$$(Y - y)(Y^{d-1} + yY^{d-2} + \cdots + y^{d-1}) = (X - x)\tilde{g}(X), \qquad \tilde{g} \in \mathbf{F}_q[X],$$

and hence

$$\frac{f}{X-x}=\sum_{i=0}^{d-1}h_iY^i+\Bigl(\sum_{i=0}^{d-1}\tilde{h}_i(Y)\Bigr)\frac{\tilde{g}(X)}{Y^{d-1}+yY^{d-2}+\cdots+y^{d-1}},$$

and this is of the desired type since, at (x,y), the last denominator standing is $dy^{d-1}\neq 0$ (remember $d\mid q-1$ and we are in the case $y\neq 0$).

In the remaining case $y=0$, we have similarly

$$\frac{f}{Y}=\frac{g_0}{Y}+\sum_{j\geqslant 1}g_iY^{i-1},$$

and

$$\frac{g_0}{Y}=\frac{g_0(x)}{Y}+\frac{X-x}{Y}\tilde{g}_0(X)=\frac{X-x}{Y}\tilde{g}_0,\qquad \tilde{g}_0\in\mathbf{F}_q[X],$$

and as above

$$\frac{X-x}{Y}=\frac{Y^{d-1}}{\tilde{g}(X)},$$

in the fraction field of $\mathcal{O}(\mathcal{C})$. This is well-defined at $\boldsymbol{x}$ since $\tilde{g}(x)=g'(x)$ and we assume g is squarefree. □

As an immediate corollary, we see the intuitively obvious fact that

$$\deg_0(fg)=\deg_0(f)+\deg_0(g). \tag{4.25}$$

Remark 4.23. The case where $y=0$ accounts for at most d points in $\mathcal{C}(\mathbf{F}_q)$. In addition, dealing with it properly in Proposition 4.20 is the only reason to impose the condition that g be squarefree (this, as we will clarify in an Appendix, has to do with issues of existence of *singularities* on the curve). For the purpose of proving Theorem 4.17, it is therefore possible to proceed without this assumption, and following the argument below, one will arrive at an upper bound

$$\begin{aligned}|\mathcal{C}(\mathbf{F}_q)|&\leqslant q+C\sqrt{q}+|\{x\in\mathbf{F}_q\mid g(x)=0\}|\\&\leqslant q+C\sqrt{q}+d\leqslant q+C'\sqrt{q}\end{aligned}$$

where $C'=d+C$ still depends only on d and $\deg(g)$. This is of course still sufficient for the Riemann Hypothesis.

Proof of Proposition 4.20. (1) The second inequality is clear since, with obvious notation, we have

$$f_1+f_2=\sum_{i=0}^{d-1}(g_{i,1}+g_{i,2})Y^i$$

and

$$d\deg(g_{i,1}+g_{i,2})+i\deg(g)\leqslant\max(d\deg(g_{i,1})\\ +i\deg(g),d\deg(g_{i_2})+i\deg(g)).$$

For the multiplicativity, fix indices i_1 and i_2 with

$$\deg(f_1)=d\deg(g_{i_1,1})+i_1\deg(g),\qquad \deg(f_2)=d\deg(g_{i_2,2})+i_2\deg(g).$$

In the product f_1f_2, there occurs a term

$$g_{i_1,1}g_{i_2,2}Y^{i_1+i_2},$$

and two cases arise: either $i_1 + i_2 \leqslant d - 1$, in which case the degree of this term is

$$d(\deg(g_{i_1,1}) + \deg(g_{i_2,2})) + (i_1 + i_2)\deg(g) = \deg(f_1) + \deg(f_2),$$

or $d - 1 < i_1 + i_2 < 2d$, in which case one rewrites the product, using the equation $Y^d = g(X)$, as

$$g_{i_1,1}g_{i_2,2}Y^{i_1+i_2} = gg_{i_1,1}g_{i_2,2}Y^{i_1+i_2-d},$$

with degree

$$d(\deg(g) + \deg(g_{i_1,1}) + \deg(g_{i_2,2})) + (i_1 + i_2 - d)\deg(g) = \deg(f_1) + \deg(f_2).$$

Hence we have $\deg(f_1 f_2) \geqslant \deg(f_1) + \deg(f_2)$. But since it is easily checked that all the other products of terms in f_1 and f_2 have strictly smaller degree, equality holds.

(2) First of all, let us see why the inequality holds when f is a "single term" $f = g_i Y^i$, $0 \leqslant i \leqslant d - 1$. In that case, the zeros of f are obtained by solving the equations

$$\begin{cases} g_i(x)y^i = 0 \\ y^d = g(x), \end{cases}$$

of which the solutions are

$$(\alpha, 0) \quad \text{where} \quad g(\alpha) = 0,$$

which counts for $\leqslant \deg(g)$ distinct solutions (or, clearly, $\leqslant i\deg(g)$ when multiplicity is taken into account), together with

$$\{(x, y) \in \bar{\mathbf{F}}_q \times \bar{\mathbf{F}}_q \mid g_i(x) = 0,\ y^d = g(x)\},$$

which has $\leqslant d\deg(g_i)$ solutions (each root of g_i giving rise to at most d solutions). Thus the total number is $\leqslant d\deg(g_i) + i\deg(g) = \deg(f)$.

We now come to the general case, and we must be careful to take multiplicity into account (since the essence of the method will be to apply this result to functions with zeros of high multiplicity). The idea is to reduce to the corresponding statement for polynomials: a non-zero $f \in \mathbf{F}_q[X]$ has at most $\deg(f)$ zeros, counted with multiplicity. To do this, denote first by K the fraction field of $\mathcal{O}(\mathcal{C})$. We first define

$$(\xi \cdot f)(x, y) = f(x, \xi y)$$

for any $f \in K$ and $\xi \in \mathbf{F}_q$ which is a d-th root of unity. Since $d \mid q - 1$, all d-th roots of unity are in $\mathbf{F}_q$, and we construct the norm map

$$(4.26) \qquad N_{\mathcal{C}} : \begin{cases} K^\times & \longrightarrow & \mathbf{F}_q(X)^\times \\ f & \mapsto & \prod_{\xi^d=1} (\xi \cdot f) \end{cases},$$

where the fact that $N_{\mathcal{C}}(f)$ depends only on the variable X is due to the fact that all $(x, y) \in \mathcal{C}$ for a given x are of the form $(x, \xi y)$, ξ running over d-th roots of unity, and by rearranging the product, the value of $N_{\mathcal{C}}(f)$ at $(x, \xi y)$ is the same as that at (x, y).

We now claim the following two facts:

(1) We have $\deg_{\mathbf{F}_q[X]} N_{\mathcal{C}}(f) = \deg(f)$, where the degree of $N_{\mathcal{C}}(f)$ is computed *in the polynomial ring* $\mathbf{F}_q[X]$;

(2) if $x \in \bar{\mathbf{F}}_q$ and y_0 is such that $(x, y_0) \in \mathcal{C}$, then assuming that for all d-th roots of unity ξ, the point $(x, \xi y_0)$ is a zero of f with multiplicity $\geqslant m_\xi(x)$, then x is a zero of the polynonial $N_{\mathcal{C}}(f)$ with multiplicity at least

$$m = \sum_{\xi^d=1} m_\xi(x).$$

If we assume these two properties, then we obtain

$$\deg_0(f) = \sum_x \sum_{\xi^d=1} m_\xi(x) \leqslant \deg_{\mathbf{F}_q[X]}(N_{\mathcal{C}}(f)) = \deg(f).$$

To check property (1), note that $\deg(\xi \cdot f) = \deg(f)$ for all ξ, and hence by (4.22), we have

$$\deg N_{\mathcal{C}}(f) = d \deg(f),$$

if the degree on the left is computed in $\mathcal{O}(\mathcal{C})$; since a polynomial $f \in \mathbf{F}_q[X] \subset \mathcal{O}(\mathcal{C})$ satisfies

$$\deg(f) = d \deg_{\mathbf{F}_q[X]}(f),$$

we obtain the equality which was claimed.

To check property (2), observe that the definition of multiplicity implies that $\xi \cdot f$ has a zero of multiplicity $\geqslant m_\xi(x)$ at the point (x, y_0). Hence

$$\frac{(\xi \cdot f)}{(X-x)^{m_\xi(x)}} = \frac{f_{1,\xi}}{f_{2,\xi}}$$

with $f_{2,\xi}(x, y_0) \neq 0$. We take the product over ξ, and we find that

$$\frac{N_{\mathcal{C}}(f)}{(X-x)^m} = \prod_{\xi^d=1} \frac{f_{1,\xi}}{f_{2,\xi}}$$

is a rational function in $\mathbf{F}_q(X)$ defined defined at x, which implies the result. □

With this done, we can come to Bombieri's construction of a suitable auxiliary function. For $k \geqslant 0$, let

$$\mathcal{H}(k) = \{f \in \mathcal{O}(\mathcal{C}) \mid \deg(f) \leqslant k\} \cup \{0\},$$

which is an $\mathbf{F}_q$-vector space, by (4.23). These spaces are analogues of the vector spaces of polynomials with a bound for the degree. The next crucial lemma shows that they behave quite similarly as k varies.

Lemma 4.24 (Riemann-Roch properties). *We have:*

(1) *The space $\mathcal{H}(0)$ is reduced to the constant functions, and for all k, we have*

$$\dim \mathcal{H}(k) \leqslant \dim \mathcal{H}(k+1) \leqslant \dim \mathcal{H}(k) + 1.$$

(2) *For all $k \geqslant 0$, we have*

$$k + 1 - (d-1)(\deg(g) - 1) \leqslant \dim \mathcal{H}(k) \leqslant k + 1,$$

and there exists $\gamma \geqslant 0$ and k_0, depending only on d and $\deg(g)$, such that

$$\dim \mathcal{H}(k) = k + 1 - \gamma$$

for all $k \geqslant k_0$. If $\gamma = 0$, then also $k_0 = 0$, and in any case one can take

$$k_0 = (d-1)(\deg(g) - 1).$$

We will explain in an Appendix below why this is a special case of the Riemann-Roch theorem, and describe briefly the latter in general.

The main property we need to prove these facts is the following very simple lemma:

Lemma 4.25. *Let $m \geqslant 1$, $d \geqslant 1$ be coprime integers. Then, for a given $k \geqslant 0$, there exists at most one pair (δ, i) of non-negative integers with $0 \leqslant i \leqslant d-1$ such that*

$$k = d\delta + im. \tag{4.27}$$

Moreover, if $k \geqslant (d-1)(m-1)$, then such a pair (δ, i) exists.

Proof. First, if $d\delta_1 + i_1 m = d\delta_2 + i_2 m$, reducing modulo d we find

$$(i_2 - i_1)m \equiv 0 \pmod{d},$$

hence $i_2 \equiv i_1 \pmod{d}$ since $(m, d) = 1$ by assumption. But then we have in fact $i_1 = i_2$ since $-d < i_2 - i_1 < d$.

To show the existence of δ and i, we argue similarly: reducing modulo d, the integer i is determined as the lift in $\{0, \ldots, d-1\}$ of $\overline{m}k \in \mathbf{Z}/d\mathbf{Z}$, where $\overline{m}$ is the inverse computed in this ring. For this index i, we have by definition

$$im - k \equiv 0 \pmod{d},$$

hence we can find $\delta \in \mathbf{Z}$ such that $d\delta + im = k$. The only issue is that δ might be negative. But note that if that is the case, it follows that

$$k = im + d\delta \leqslant im - d \leqslant (d-1)m - d < (d-1)(m-1),$$

hence the result by contraposition. □

Proof of Lemma 4.24. That $\mathcal{H}(0)$ is the space of constant functions is clear, and obviously $\mathcal{H}(k) \subset \mathcal{H}(k+1)$ so the dimensions are non-decreasing. Now consider two elements f_1, $f_2 \in \mathcal{H}(k+1)$, neither of which is in $\mathcal{H}(k)$. Writing

$$f_1 = \sum_{i=0}^{d-1} g_{i,1} Y^i, \qquad f_2 = \sum_{i=0}^{d-1} g_{i,2} Y^i,$$

if follows from Lemma 4.25 applied with $m = \deg(g)$ that there is a single index i such that

$$k + 1 = \deg(f_1) = d \deg(g_{i,1}) + i \deg(g),$$
$$k + 1 = \deg(f_2) = d \deg(g_{i,2}) + i \deg(g).$$

In particular, $g_{i,1}$, $g_{i,2}$ have the same degree, say δ, and hence there exist α, $\beta \in \mathbf{F}_q$ such that

$$\deg(\alpha g_{i,1} + \beta g_{i,2}) < \delta,$$

and it follows that

$$\deg(\alpha f_1 + \beta f_2) < \deg(f_1) = k + 1,$$

so that we have shown that $\dim \mathcal{H}(k+1)/\mathcal{H}(k) \leqslant 1$.

From this last inequality and $\dim \mathcal{H}(0) = 1$, the upper bound

$$\dim \mathcal{H}(k) \leqslant k + 1$$

is clear by induction. For the lower bound, we will show that for all

$$k \geqslant k_0 = (d-1)(\deg(g) - 1),$$

there exists an element $0 \neq f \in \mathcal{O}(\mathcal{C})$ of degree exactly k in $\mathcal{H}(k)$. This will clearly show that

$$\dim \mathcal{H}(k) \geqslant k - k_0 + \dim \mathcal{H}(0) = k + 1 - (d-1)(\deg(g) - 1),$$

and in fact, using (1), that for $k \geqslant k_0$ we have

$$\dim \mathcal{H}(k) = \dim \mathcal{H}(k_0) + (k - k_0) = k + 1 - \gamma$$

with

$$\gamma = k_0 - \dim \mathcal{H}(k_0) + 1 \leqslant k_0 + 1. \tag{4.28}$$

Note that $\gamma = 0$ means that $\dim \mathcal{H}(k_0) = k_0 + 1$, which is only possible if $\dim \mathcal{H}(k) = k + 1$ also for $k \leqslant k_0$, i.e., if one can take $k_0 = 0$.

Now, to construct f of degree k, it is enough to show that we can find integers (δ, i) with $0 \leqslant \delta$, $0 \leqslant i \leqslant d-1$, such that

$$d\delta + i\deg(g) = k,$$

since we can then take $f = X^{\delta}Y^{i}$. Thus the second part of Lemma 4.25 gives the result. □

Finally, we can describe Bombieri's construction. First let

$$0 = d_0 < d_1 < \cdots < d_n < \cdots$$

be the increasing sequence of indices such that

$$\dim \mathcal{H}(d_j + 1) > \dim \mathcal{H}(d_j),$$

and let $s_j \in \mathcal{H}(d_j)$ be an element not in the previous space $\mathcal{H}(d_j - 1)$ (these are, in some sense, analogues of the basis X^j of a polynomial ring $K[X]$, where d_j would be j).

Now denote

$$S_j = s_j \circ \mathrm{Fr},$$

which is still an element of $\mathcal{O}(\mathcal{C})$. We look for auxiliary functions of the type

$$f = \sum_j f_j^m S_j,$$

where the sum is of course assumed to be finite and where $m = p^a$ for some a, $0 \leqslant a \leqslant [\mathbf{F}_q : \mathbf{F}_p]$. Our first remark is quite easy:

Lemma 4.26. *For $\kappa \geqslant 0$, the set of functions*

$$f = \sum_j f_j^m S_j, \tag{4.29}$$

where j runs over integers such that $0 \leqslant d_j \leqslant \kappa$ and where $f_j \in \mathcal{H}(k)$, is an $\mathbf{F}_q$-vector space, denoted $\tilde{\mathcal{H}}(m,k,\kappa)$, which is equal to the vector space spanned by products of the type $f_1^m \cdot (f_2 \circ \mathrm{Fr})$ where $f_1 \in \mathcal{H}(k)$ and $f_2 \in \mathcal{H}(\kappa)$.

Proof. Let first $\tilde{\mathcal{H}}(m,k,\kappa)$ denote the vector space generated by the products described. Clearly any function f of the form (4.29) is in $\tilde{\mathcal{H}}(m,k,\kappa)$, and it is enough to show conversely that any $f \in \tilde{\mathcal{H}}(m,k,\kappa)$ can be written in this form.

We first do this for $f = f_1^m \cdot (f_2 \circ \mathrm{Fr})$. The functions s_j, where $d_j \leqslant \kappa$, form a basis of $\mathcal{H}(\kappa)$, and hence we can write

$$f_2 = \sum_j \alpha_j s_j, \qquad f_2 \circ \mathrm{Fr} = \sum_j \alpha_j S_j,$$

for some $\alpha_j \in \mathbf{F}_q$. Since $\alpha_j = \alpha_j^q = (\alpha_j^{q/m})^m$ (recall that $m \mid q$), we can also write

$$f_1^m (f_2 \circ \mathrm{Fr}) = \sum_j (\alpha_j^{q/m} f_1)^m S_j,$$

which is indeed of type (4.29).

Now given α, $\beta \in \mathbf{F}_q$, and

$$f = \sum_j f_j^m S_j, \quad h = \sum_j h_j^m S_j$$

we write again $\alpha = (\alpha^{q/m})^m$, $\beta = (\beta^{q/m})^m$ as above, so that

$$\alpha f + \beta h = \sum_j (\alpha^{q/m} f_j + \beta^{q/m} h_j)^m S_j,$$

using the additivity of the p-th power operation in characteristic p. □

This type of auxiliary functions is potentially useful because of the following other simple fact:

Lemma 4.27. *Let f be any function of the type (4.29). Then, if*

$$\sum_j f_j^m s_j = 0 \in \mathcal{O}(\mathcal{C}), \tag{4.30}$$

it follows that for any $\boldsymbol{x} \in \mathcal{C}(\mathbf{F}_q)$, we have $f(\boldsymbol{x}) = 0$, and in fact f vanishes to order at least $m = p^a$ at $\boldsymbol{x}$.

Proof. First of all, we compute $f(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathcal{C}(\mathbf{F}_q)$: we have

$$\begin{aligned} f(\boldsymbol{x}) &= \sum_j f_j^m(\boldsymbol{x}) S_j(\boldsymbol{x}) \\ &= \sum_j f_j^m(\boldsymbol{x}) s_j(\mathrm{Fr}(\boldsymbol{x})) \\ &= \sum_j f_j^m(\boldsymbol{x}) s_j(\boldsymbol{x}) = 0, \end{aligned}$$

by (4.30) *and the crucial identity* $\mathrm{Fr}(\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in \mathcal{C}(\mathbf{F}_q)$. Furthermore, by definition of the Frobenius, we have

$$S_j = s_j \circ \mathrm{Fr} = s_j^q,$$

which is an m-th power in $\mathcal{O}(\mathcal{C})$ since $m = p^a$ with $p^a \mid q$, and f_j^m is also an m-th power. Hence (again because m is a power of p), the function f is an m-th power in $\mathcal{O}(\mathcal{C})$, and it follows that each of its zero has multiplicity $\geqslant m$. □

The question is now to construct a specific auxiliary function of small degree; in particular, it must be *non-zero*. For this, which is really the fundamental step, we have the following lemma:

Lemma 4.28. *Let f be an auxiliary function of the type*

$$f = \sum_j f_j^m S_j,$$

with $f_j \in \mathcal{H}(k)$ for all j and $m = p^a \mid q$. Then, provided

$$km < q$$

we have $f = 0$ if and only if $f_j = 0$ for all j. In other words, for any $\kappa \geqslant 0$, the representation (4.29) *of a function $f \in \tilde{\mathcal{H}}(m, k, \kappa)$ is* unique.

Proof. Suppose the (f_j) are not all zero but f is zero. Let then $i \geqslant 0$ be such that $f_i \neq 0$ but $f_j = 0$ for $j > i$, so that we have

$$f_i^m S_i = -\sum_{j<i} f_j^m S_j.$$

The degree of the left-hand side is given (by (4.22)) by

$$m \deg(f_i) + q \deg(s_i) = m \deg(f_i) + q d_i,$$

while, by (4.23), we have

$$\deg\Bigl(\sum_{j>i} f_j^m S_j\Bigr) \leqslant \max\{m\deg(f_j)+qd_j \mid j<i\} \leqslant mk+qd_{i-1}$$

since the sequence of indices d_j is increasing. Thus the equality means that

$$q(d_i-d_{i-1}) \leqslant m(k-\deg(f_i)) \leqslant mk,$$

and since $d_i > d_{i-1}$, we get

$$q \leqslant mk.$$

The lemma is thus proved by contraposition. □

Now, in addition to having $f \neq 0$, we must ensure the condition (4.30) for f with small degree. But if we restrict to f of the type (4.29) with $d_j \leqslant \kappa$, this condition can be written $\Delta(f)=0$ for the $\mathbf{F}_p$-linear map

$$\Delta : \begin{cases} \tilde{\mathcal{H}}(m,k,\kappa) & \longrightarrow \quad \mathcal{H}(mk+\kappa) \\ \displaystyle\sum_j f_j^m S_j & \mapsto \quad \displaystyle\sum_j f_j^m s_j \end{cases}.$$

The fact that Δ is well-defined (and its $\mathbf{F}_p$-linearity; note it is *not* $\mathbf{F}_q$-linear) follows from Lemma 4.28, and from the fact that

$$\deg(f_j^m s_j) \leqslant mk+\kappa.$$

Now by the last part of Lemma 4.24, we have

$$\begin{aligned}\dim_{\mathbf{F}_p}\operatorname{Ker}(\Delta) &\geqslant \dim_{\mathbf{F}_p}\tilde{\mathcal{H}}(m,k,\kappa)-\dim_{\mathbf{F}_p}\mathcal{H}(mk+\kappa)\\ &= [\mathbf{F}_q:\mathbf{F}_p]\{(\dim\mathcal{H}(k))(\dim\mathcal{H}(\kappa))-\dim\mathcal{H}(mk+\kappa)\}\\ &= [\mathbf{F}_q:\mathbf{F}_p]\{(k+1-\gamma)(\kappa+1-\gamma)-(mk+\kappa+1-\gamma)\},\end{aligned}$$

for k, $m \geqslant k_0=(d-1)(\deg(g)-1)$, where $\gamma \leqslant (d-1)(\deg(g)-1)$ also (by (4.28)), and we used again Lemma 4.28 to compute the dimension of $\tilde{\mathcal{H}}(m,k,\kappa)$.

The inescapable conclusion is the following: provided

$$\begin{gathered} km<q,\ \ k\geqslant k_0,\ \ m\geqslant k_0, \\ (k+1-\gamma)(\kappa+1-\gamma)-(mk+\kappa+1-\gamma)>0, \end{gathered} \tag{4.31}$$

we can find $f \neq 0$ in $\operatorname{Ker}(\Delta)$, and then we know that

$$|\mathcal{C}(\mathbf{F}_q)| \leqslant \frac{\deg(f)}{m} \leqslant \frac{mk+q\kappa}{m} = k+\frac{q\kappa}{m}.$$

We must now optimize these parameters. If we compare the upper bound with the goal of Theorem 4.17, we must clearly take κ and m very nearly equal, say

$$\kappa+1-\gamma = m+C$$

for some $C \geqslant 0$. The upper bound becomes

$$|\mathcal{C}(\mathbf{F}_q)| \leqslant q+k+\frac{(C-1+\gamma)q}{m}$$

and to minimize the "error term", k and m should also be very close, and indeed close to $\sqrt{q}$. Since m is a power of a prime, this explains why we assumed that $q=p^{2\nu}$ for some $\nu \geqslant 1$, as it allows us to take $m=p^\nu=\sqrt{q}$.

Now the last constraint (4.31) (which ensures the existence of the auxiliary function) translates easily to

$$k > \Bigl(1+\frac{m}{C}\Bigr)\gamma$$

and the condition $km = k\sqrt{q} < q$ shows that C should be $> \gamma$. We take in fact $C = \gamma+1$, so that

$$\kappa = \sqrt{q} + 2\gamma,$$

and then the condition on k is

$$k > \frac{\gamma}{\gamma+1}m + \gamma,$$

which is true for

$$k = \Bigl\lfloor \frac{\gamma}{\gamma+1}m \Bigr\rfloor + \gamma + 1.$$

The condition $km = k\sqrt{q} < q$ is then satisfied if

$$m\Bigl(\frac{\gamma}{\gamma+1}m + \gamma + 1\Bigr) < q \;\Leftrightarrow\; q > (\gamma+1)^4,$$

and $k \geqslant k_0$, $m \geqslant k_0$ will be satisfied (at least) for $q > (2k_0)^2$ (to check this, the case $\gamma = 0$ must be treated separately by noticing that we have $k_0 = 0$ in that case). Hence, for $q > q_0 = \max((\gamma+1)^4, k_0^2)$, we obtain

$$|\mathcal{C}(\mathbf{F}_q)| \leqslant q + 2\gamma\sqrt{q} + k \leqslant q + (2\gamma+1)\sqrt{q}, \qquad \text{since } k < qm^{-1} = \sqrt{q}.$$

We take care of small values of q as usual by writing

$$|\mathcal{C}(\mathbf{F}_q)| \leqslant q + d\sqrt{q_0}\sqrt{q}$$

in all cases, since the trivial bound is

$$|\mathcal{C}(\mathbf{F}_q)| = |\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^d = g(x)\}| \leqslant dq$$

which is $\leqslant d\sqrt{q_0}\sqrt{q}$ if $q \leqslant q_0$.

Remark 4.29. As a special case of Theorem 4.17, if q is odd and $g \in \mathbf{F}_q[X]$ is monic squarefree of degree 3, say

$$g = X^3 + aX + b,$$

we obtain

$$|\{(x,y) \in \mathbf{F}_q \times \mathbf{F}_q \mid y^2 = x^3 + ax + b\}| = q - a$$

where

$$|a| \leqslant 2\sqrt{q}.$$

This statement was the first (non-trivial) instance of the Riemann Hypothesis to be proved, by H. Hasse. His first proof was very different.

Appendix: Hasse derivatives

The idea of Hasse derivatives is to “eliminate” the factor $p!$ from the p-th derivative by a clever formal construction.

Definition 4.30 (Hasse derivative). Let k be a field and $j \geqslant 0$ an integer. The *j-th Hasse derivative* is the k-linear map

$$\begin{cases} k[X] & \longrightarrow & k[X] \\ X^n & \mapsto & \dbinom{n}{j} X^{n-j} \end{cases}$$

with the convention that $X^n \mapsto 0$ if $j > n$. We denote $f^{[j]}$ the j-th Hasse derivative of $f \in k[X]$, and note that $f^{[0]} = f$.

Note that the j-th derivative is given by

$$X^n \mapsto n(n-1)\cdots(n-j+1)X^{n-j} = j!\binom{n}{j}X^{n-j},$$

hence we have by linearity

$$f^{(j)} = j!f^{[j]}$$

for all $f \in k[X]$.

Thus, if $j!$ is invertible in k, there is not much difference between the classical and the Hasse derivatives. But in characteristic p, this only happens for $j < p$, and for $j \geqslant p$, although the derivatives vanish entirely, the Hasse derivatives do not. Indeed, we have

$$(X^p)^{[p]} = \binom{p}{p} = 1 \neq 0.$$

In particular, for this particular polynomial, observing the vanishing of Hasse derivatives gives the order of vanishing at 0. This is a general fact.

Proposition 4.31 (Properties of Hasse derivatives). *Let k be a field.*

(1) *For all $f_1, \ldots, f_r \in k[X]$ and $j \geqslant 0$ we have*

$$(f_1\cdots f_r)^{[j]} = \sum\cdots\sum_{j_1+\cdots+j_r=j} f_1^{[j_1]}\cdots f_r^{[j_r]},$$

in particular

$$(f_1f_2)^{[j]} = \sum_{i=0}^{j} f_1^{[i]}f_2^{[j-i]}.$$

(2) *Let $a \in k$ be given. Then for $f \in k[X]$, we have*

$$f = \sum_{j=0}^{\deg(f)} f^{[j]}(a)(X-a)^j. \tag{4.32}$$

(3) *An element $a \in k$ is a zero of order $\geqslant m \geqslant 1$ of a polynomial $f \in k[X]$ if and only if*

$$f(a) = f^{[1]}(a) = \cdots = f^{[m-1]}(a) = 0.$$

Proof. (1) The general case follows from the special case $r = 2$ by induction on r. For the latter, it is enough by linearity to consider $f_1 = X^{d_1}$, $f_2 = X^{d_2}$, and then the identity is equivalent with the formula

$$\binom{d_1+d_2}{j} = \sum_{i=0}^{j}\binom{d_1}{i}\binom{d_2}{j-i}.$$

for binomial coefficients, which can be proved, e.g., by comparing the coefficients of X^j on both sides of the obvious identity

$$(X+1)^{d_1+d_2} = (X+1)^{d_1}(X+1)^{d_2}.$$

(2) The identity says that the values of the Hasse derivatives at a are the same as the coefficients of the Taylor expansion of f in powers of $(X-a)$. To show this, since the map sending f to the coefficient of $(X-a)^j$ in the basis of powers of $X-a$ is itself

linear, we need only compute the Hasse derivatives of $h = (X-a)^d$ for all $d \geqslant 0$. Now, by the binomial expansion, we compute

$$\begin{aligned} h^{[j]} &= \Bigl(\sum_{i=0}^{d} \binom{d}{i} (-a)^{d-i} X^i\Bigr)^{[j]} = \sum_{i=0}^{d} \binom{d}{i} (-a)^{d-i} \binom{i}{j} X^{i-j} \\ &= \sum_{\iota=0}^{d} \binom{d}{\iota} (-a)^{\iota} \binom{d-\iota}{j} X^{d-i-j} \\ &= \binom{d}{j} \sum_{\iota=0}^{d} \binom{d-j}{\iota} (-a)^{\iota} X^{d-i-j} = \binom{d}{j} (X-a)^{d-j}, \end{aligned} \tag{4.33}$$

for $j \geqslant 0$; we have used the identity

$$\binom{d}{\iota}\binom{d-\iota}{j} = \frac{d!}{\iota! j! (d-j-\iota)!} = \binom{d}{j}\binom{d-j}{\iota}.$$

This relation gives (4.32) in the case of h.

(3) Assume first that a is a zero of order $\geqslant \ell$ of f, so that

$$f = (X-a)^{\ell} g = hg, \qquad \text{with } h = (X-a)^{\ell},$$

for some polynomial $g \in k[X]$. By (1), we have

$$f^{[j]} = \sum_{i=0}^{j} h^{[i]} g^{[j-i]}$$

for any $j \geqslant 0$, and hence

$$f^{[j]}(a) = h(a) g^{[j]}(a) + \cdots + h^{[j]}(a) g(a),$$

so that it is enough to show that $h^{[j]}(a) = 0$ for $j < \ell$. (In other words, we have reduced the problem to the case of h). But (4.33) gives immediately $h^{[j]}(a) = 0$ if $j < \ell$.

Now for the converse, if, for some $\ell \geqslant 1$, we have

$$f(a) = f^{[1]}(a) = \cdots = f^{[\ell-1]}(a) = 0,$$

the relation (4.32) shows that

$$\begin{aligned} f &= \sum_{j=0}^{\deg(f)} f^{[j]}(a)(X-a)^j = \sum_{j=\ell}^{\deg(g)} f^{[j]}(a)(X-a)^j \\ &= (X-a)^{\ell} \sum_{j=\ell}^{\deg(g)} f^{[j]}(a)(X-a)^{j-\ell} \end{aligned}$$

is divisible by $(X-a)^{\ell}$ in $k[X]$, as desired. □

CHAPTER 5

Additive character sums

We now come to the proof of Theorem 3.2, involving additive character sums. In some sense, the proof is very similar to the proof of the Riemann Hypothesis for multiplicative character sums. However, the similarity comes in spite of very striking differences between the two situations, which we will also emphasize throughout.

We will use information from the successful proof of Theorem 3.1 to motivate the steps we are going to take, but these will not be in the same order as before.

Indeed, we used three main ingredients in the previous chapter:

- The introduction of a suitable L-function (associated to a Dirichlet character) to represent the exponential sum and its relatives over extension fields as a finite sum (with at most $d-1$ terms) of powers of some fixed complex numbers;
- An averaging trick to reduce the estimation of a family of these sums of ω_j^ν to a point counting problem;
- The use of Bombieri's version of the Stepanov method to obtain very good bounds for the point counting.

It turns out that the second and third steps can be adapted very easily to additive character sums. The first one is where the main difference lies, as it is not possible to obtain an L-function for these sums using only Dirichlet characters, and one needs a generalization of these. Because of this, we first cover the last steps in the next two sections, before dealing with L-functions for additive characters.

5.1. Reduction to point counting

Consider once more a finite field $\mathbf{F}_q$ with q elements and the ring $\mathbf{F}_q[X]$ of polynomials in one variable with coefficients in $\mathbf{F}_q$. For a fixed $g \in \mathbf{F}_q[X]$, and any additive character

$$\psi : \mathbf{F}_q \to \mathbf{C}^\times,$$

we want to bound the exponential sum

$$\sum_{x \in \mathbf{F}_q} \psi(g(x)).$$

It is natural to expect that, in addition to the sum over $\mathbf{F}_q$, the sums

$$S_\nu(g,\psi) = \sum_{x \in \mathbf{F}_{q^\nu}} \psi(\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))), \tag{5.1}$$

ranging over extensions $\mathbf{F}_{q^\nu}$ of $\mathbf{F}_q$, will play a role.

Indeed, the easy first reduction is to go from bounding character sums to bounding the number of points on some "curve" over finite fields. The analogue of Lemma 4.16 is the following:

Lemma 5.1 (Reduction to point counting). *Let $\mathbf{F}_q$ be a finite field, g a non-constant monic polynomial in $\mathbf{F}_q[X]$, ψ a fixed non-trivial additive character of $\mathbf{F}_q$. For all $\nu \geqslant 1$,* *we have*

$$\sum_{a\in\mathbf{F}_q^\times}\sum_{x\in\mathbf{F}_{q^\nu}}\psi(\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(ag(x))) = \\ |\{(x,y)\in\mathbf{F}_{q^\nu}\times\mathbf{F}_{q^\nu}\ \mid\ y^q-y=g(x)\}|-q^\nu.$$

Proof. This is a simple consequence of Lemma 1.3: since the characters

$$x\mapsto\psi(ax),$$

where $a\in\mathbf{F}_q$, are all the additive characters of $\mathbf{F}_q$, we have

$$\begin{aligned}\sum_{a\in\mathbf{F}_q}\sum_{x\in\mathbf{F}_{q^\nu}}\psi(a\,\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))) &= \sum_{x\in\mathbf{F}_{q^\nu}}\sum_{a\in\mathbf{F}_q}\psi(a\,\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x)))\\ &= q|\{x\in\mathbf{F}_{q^\nu}\ \mid\ \mathrm{Tr}(g(x))=0\}|\end{aligned}$$

by orthogonality of characters. But $\mathrm{Tr}(g(x))=0$ if and only if there exists $y\in\mathbf{F}_{q^\nu}$ such that $y^q-y=g(x)$, by Lemma 1.3. Moreover, if that is the case, there are q solutions to the equation

$$y^q-y=g(x),$$

which are given by $y=y_0+\alpha$ with $\alpha\in\mathbf{F}_q$, in terms of a fixed solution y_0 (indeed, we must have $(y-y_0)^q=(y-y_0)$ so $y-y_0\in\mathbf{F}_q$), so that

$$q|\{x\in\mathbf{F}_{q^\nu}\ \mid\ \mathrm{Tr}(g(x))=0\}|=|\{(x,y)\in\mathbf{F}_{q^\nu}\times\mathbf{F}_{q^\nu}\ \mid\ y^q-y=g(x)\}|.$$

This corresponds to the sum over all $a\in\mathbf{F}_q$; the contribution of $a=0$ is of course q^ν, so that we obtain the statement after subtraction. □

5.2. Implementing Bombieri's method

What remains to do now is to prove the following point-counting estimate, which is the analogue of Theorem 4.17 in the previous chapter:

Theorem 5.2. *Let $\mathbf{F}_q$ be a finite field, $g\in\mathbf{F}_q[X]$ a non-constant monic polynomial of degree d with $d<q$ and $(d,q)=1$. Then there exists a constant $C\geqslant 0$, depending only on d and q, such that for al $\nu\geqslant 1$, we have*

$$||\{(x,y)\in\mathbf{F}_{q^{2\nu}}\times\mathbf{F}_{q^{2\nu}}\ \mid\ y^q-y=g(x)\}|-q^{2\nu}|\leqslant Cq^\nu.$$

Note that this statement, for $\nu=1$, is trivial if the constant C depends on d and q (as it does), so we think of it as an asymptotic result.

We want to proceed as in Section 4.3, and this turns out to be surprisingly easy. We first have the analogue of Lemma 4.18, which allows us to reduce the two-sided estimate to an upper-bound only, accessible by means of the Stepanov method.

Lemma 5.3 (From upper bound to lower bound). *Let $\mathbf{F}_q$ be a finite field. For $\nu\geqslant 1$ and polynomials $g\in\mathbf{F}_{q^\nu}[X]$ with degree $d\geqslant 1$ with $(d,p)=1$ and $d<q$, let $a(g)\in\mathbf{Z}$ be defined by the equality*

$$|\{(x,y)\in\mathbf{F}_{q^\nu}\times\mathbf{F}_{q^\nu}\ \mid\ y^q-y=g(x)\}|=q^\nu+a(g).$$

Then we have

$$|\{(x,y)\in\mathbf{F}_{q^\nu}\times\mathbf{F}_{q^\nu}\ \mid\ y^q-y=g(x)\}|\geqslant q^\nu-(q-1)\max_{\alpha\in\mathbf{F}_{q^\nu}}|a(g+\alpha)|.$$

Proof. This is the same type of argument used in Lemma 4.18, where intead of the cyclic group $\mathbf{F}_q^\times/(\mathbf{F}_q^\times)^d$, we use the (additive) group

$$AS = \mathbf{F}_{q^\nu}/\langle y^q - y \mid y \in \mathbf{F}_{q^\nu}\rangle$$

of order q. Let

$$\{0, \alpha_2, \ldots, \alpha_{q-1}\}$$

be representatives of AS in $\mathbf{F}_{q^\nu}$. For any $x \in \mathbf{F}_{q^\nu}$, $g(x)$ is of the form $y^q - y + \alpha$ for some unique α in this set of representatives, and for q different values of y. Thus using the auxiliary polynomials $h_\alpha = g - \alpha$ and

$$\mathcal{X}_\alpha = \{(x, y) \in \mathbf{F}_{q^\nu} \times \mathbf{F}_{q^\nu} \mid y^q - y = h_\alpha(x)\},$$

we obtain the stated result by looking at the formula

$$q \cdot q^\nu = \sum_\alpha |\mathcal{X}_\alpha| = q \cdot q^\nu + \sum_\alpha a(g_\alpha).$$

□

From now on, we proceed to show that, for $g \in \mathbf{F}_{q^{2\nu}}[X]$ of degree d with $(d, p) = 1$, $d < q$, we have

(5.2) $$|\{(x, y) \in \mathbf{F}_{q^{2\nu}} \times \mathbf{F}_{q^{2\nu}} \mid y^q - y = g(x)\}| \leqslant q^{2\nu} + Dq^\nu$$

where C depends only on d and q.

For simplicity of notation, we let

$$\mathbf{F} = \mathbf{F}_{q^\nu},$$

and consider $g \in \mathbf{F}[X]$ of degree coprime with p and $< q$. Only at the end will it be needed to assume ν is even. We are now looking at the curve

$$\mathcal{X} = \{(x, y) \in \bar{\mathbf{F}}_q \times \bar{\mathbf{F}}_q \mid y^q - y = g(x)\},$$

and want to estimate the cardinality of the set $\mathcal{X}(\mathbf{F})$ of points with coordinates in $\mathbf{F}$. We will do this by finding an auxiliary function f on $\mathcal{X}$ vanishing to high order at all points in $\mathcal{X}(\mathbf{F})$ and with "small degree". Here are the definitions to proceed with Bombieri's construction:

Definition 5.4 (Degree and multiplicity for $\mathcal{X}$). Let $\mathbf{F}_q$ be a finite field with q elements, $\nu \geqslant 1$, $\mathbf{F} = \mathbf{F}_{q^\nu}$, and let $g \in \mathbf{F}[X]$ be a non-constant polynomial of degree d with $(d, q) = 1$, $d < q$. Let

$$\mathcal{O}(\mathcal{X}) = \mathbf{F}[X, Y]/(Y^q - Y - g)$$

be the integral ring of functions on $\mathcal{X}$, and $K(\mathcal{X})$ its quotient field.

(1) Any $f \in \mathcal{O}(\mathcal{X})$ can be expressed in a unique way as

$$f = \sum_{j=0}^{q-1} g_j Y^j$$

with $g_j \in \mathbf{F}[X]$, and if $f \neq 0$, the *degree of* f is defined by

$$\deg(f) = \max\{q \deg(g_i) + id \mid 0 \leqslant i \leqslant q - 1\} \geqslant 0,$$

(2) If $\boldsymbol{x} \in \mathcal{X}$, $f \in \mathcal{O}(\mathcal{X})$ and $m \geqslant 1$, one says that f vanishes to order $\geqslant m$ at $\boldsymbol{x}$ if and only if there exist f_1, $f_2 \in \mathcal{O}(\mathcal{X})$ such that $f_2(\boldsymbol{x}) \neq 0$ and

$$\frac{f}{X - x} = \frac{f_1}{f_2} \in K(\mathcal{X}).$$

Remark 5.5. With the correspondance $d \leftrightarrow q$, $\deg(g) \leftrightarrow d$, the definition of the degree is identical with that used for the curves $\mathcal{C}$ that appeared in Section 4.3.

The definition of the multiplicity is simpler because there is no need to distinguish between two types of points to define it: the function $X - x$ can always be used to determine the multiplicity. The reason is the following: for any $z \in \mathbf{F}$ (for instance, $z = g(x)$) the equation

$$y^q - y = z$$

has exactly q distinct roots in an algebraic closure of $\mathbf{F}$ (because the derivative in $\mathbf{F}[X]$ of the relevant polynomial is the constant -1, hence has no common zero with $Y^q - Y - z$), whereas, for $d \mid q-1$, the equations

$$y^d = z$$

may have a single multiple root, in the special case $y = z = 0$.

We then have the exact analogue of Proposition 4.20:

Proposition 5.6. *Let $\mathbf{F}_q$ be a finite field with q elements, $\nu \geqslant 1$, $\mathbf{F} = \mathbf{F}_{q^\nu}$ and let $g \in \mathbf{F}[X]$ be a non-constant polynomial of degree d with $(d, q) = 1$, $d < q$.*

(1) *We have*

$$\deg(f_1 f_2) = \deg(f_1) + \deg(f_2),$$
$$\deg(f_1 + f_2) \leqslant \max(\deg(f_1), \deg(f_2)),$$

for any f_1, $f_2 \in \mathcal{O}(\mathcal{X})$, with $f_1 + f_2 \neq 0$ in the second case.

(2) *For $f \in \mathcal{O}(\mathcal{X})$, $f \neq 0$, the number of zeros of f with multiplicity is at most the degree* $\deg(f)$ *of f.*

Proof. The proof of (1) is left as an exercise: it is very similar to the corresponding part of Proposition 4.20. The proof of (2) is also essentially identical, after one introduces

$$(a \cdot f)(x, y) = f(x, y + a)$$

for $a \in \mathbf{F}_q$ and the norm map

$$N_{\mathcal{X}} : K(\mathcal{X})^\times \to \mathbf{F}(X)^\times$$

by

$$N_{\mathcal{X}} f(x, y) = \prod_{a \in \mathbf{F}_q} f(x, y + a),$$

(which can be seen as identical as the definition in (4.26) after noticing that in both cases the value of the norm of f at $x \in \bar{\mathbf{F}}_q$ is the product of the values of f over all points $\boldsymbol{x}$ in the curve with first coordinate x). It is easy to see that $N_{\mathcal{X}}$ does map $\mathcal{O}(\mathcal{X})$ to $\mathbf{F}[X]$, and proceeding as in Proposition 4.20, one finds

$$\deg_{\mathbf{F}[X]}(N_{\mathcal{X}}(f)) = \deg_{\mathcal{O}(\mathcal{X})}(f),$$

from which the remainder of the argument is almost identical with the earlier one. □

One can now define "Riemann-Roch" spaces as before:

$$\mathcal{H}(k) = \{0\} \cup \{f \in \mathcal{O}(\mathcal{X}) \mid \deg(f) \leqslant k\},$$

and the analogue of Lemma 4.24 holds:

Lemma 5.7 (Riemann-Roch properties). *With notation as before, we have:*
(1) *The space $\mathcal{H}(0)$ is reduced to the constant functions, and for all k, we have*

$$\dim \mathcal{H}(k) \leqslant \dim \mathcal{H}(k+1) \leqslant \dim \mathcal{H}(k) + 1.$$

(2) *For all $k \geqslant 0$, we have*

$$k + 1 - (d-1)(\deg(g) - 1) \leqslant \dim \mathcal{H}(k) \leqslant k + 1,$$

and there exists $\gamma \geqslant 0$ and k_0, depending only on d and q, such that

$$\dim \mathcal{H}(k) = k + 1 - \gamma$$

for all $k \geqslant k_0$. If $\gamma = 0$, then also $k_0 = 0$, and in any case one can take

$$k_0 = (d-1)(q-1).$$

In view of the definition of the degree, this follows from Lemma 4.25 applied with $(d, m) = (q, d)$ and the same arguments used in Proposition 4.24.

One can then follow the remainder of Bombieri's construction, with the data denoted $(d, \deg(g), q)$ there replaced by $(q, d, q^{2\nu})$, and the conclusion of Theorem 5.2 comes out.

Remark 5.8. Since Bombieri's original paper [1] gives a uniform proof of point-counting estimates for arbitrary (non-singular) curves over finite fields, it is not surprising that the versions we presented work also (almost) identically for the curves $\mathcal{C}$ of the previous chapter and $\mathcal{X}$ in this section.

5.3. The L-functions associated to additive character sums

The last step is to construct a suitable L-function for additive character sums. This requires more extensive changes to the framework of the previous chapter because, as it turns out, one can not construct *Dirichlet characters* of $\mathbf{F}_q[X]$ to recover additive exponential sums. One needs to extend slightly the objects to which L-functions are attached.

To explain the construction, we may start using as motivation Proposition 4.7. This gives an a priori description of the L-function from character sums, and suggests that, given ψ and g, we define

$$Z(g, \psi; T) = \exp\Bigl(\sum_{\nu \geqslant 1} \frac{S_\nu(g, \psi)}{\nu} T^\nu\Bigr) \tag{5.3}$$

where $S_\nu(g, \psi)$ is given by (5.1). This expression makes sense in the ring $\mathbf{C}[[T]]$ of formal power series. Then it is enough to prove the following to conclude successfully the proof of Theorem 3.2:

Proposition 5.9 (Rationality of L-function). *Let $\mathbf{F}_q$ be a finite field, ψ a non-trivial additive character of $\mathbf{F}_q$ and $g \in \mathbf{F}_q[X]$ a non-constant polynomial of degree d such that $(d, q) = 1$. The L-function $Z(g, \psi; T)$ given by (5.3) is a polynomial in $\mathbf{C}[T]$ of degree $\leqslant (d-1)$.*

A first step towards this result is to see that this L-function function $Z(g, \psi; T)$ still retains one feature of those associated with Dirichlet characters: an Euler product expansion over irreducible monic polynomials.

Lemma 5.10 (Euler product). *Let $\mathbf{F}_q$ be a finite field, ψ a non-trivial additive character of $\mathbf{F}_q$ and $g \in \mathbf{F}_q[X]$ a non-constant polynomial of degree $d < q$ such that $(d, q) = 1$. Then we have*

$$Z(g, \psi; T) = \prod_{\pi} (1 - \eta(\pi) T^{\deg(\pi)})^{-1}$$

where π runs over irreducible monic polynomials in $\mathbf{F}_q[X]$ and

$$\eta(\pi) = \psi(\mathrm{Tr}_{\mathbf{F}_{q^{\deg(\pi)}}/\mathbf{F}_q}(g(\alpha)))$$

for any root α of π in $\bar{\mathbf{F}}_q$.

Note that since $\eta(\pi)$ depends only on the trace of the root α, it is independent of the latter. Also, despite the similarity in notation with the previous chapter, we emphasize again that η is *not* a Dirichlet character.

Proof. This is roughly the proof of Proposition 4.7, run backwards. Indeed, we have first

$$T \frac{Z'}{Z}(g, \psi; T) = \sum_{\nu \geqslant 1} S_\nu(g, \psi) T^\nu.$$

On the other hand, we can rewrite the sum over $x \in \mathbf{F}_{q^\nu}$ defining $S_\nu(g, \psi)$ in terms of the minimal polynomials π_x of those elements, and their roots:

$$\begin{aligned}
S_\nu(g, \psi) &= \sum_{x \in \mathbf{F}_{q^\nu}} \psi(\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}(g(x))) \\
&= \sum_{\deg(\pi) \mid \nu} \sum_{\substack{x \in \mathbf{F}_{q^\nu} \\ \pi(x)=0}} \psi(\mathrm{Tr}_{\mathbf{F}_{q^\nu}/\mathbf{F}_q}\, g(x)) \\
&= \sum_{d \mid \nu} \sum_{\deg(\pi)=d} \sum_{\pi(x)=0} \psi\Big(\frac{\nu}{d} \mathrm{Tr}_{\mathbf{F}_{q^d}/\mathbf{F}_q}(g(x))\Big) \\
&= \sum_{d \mid \nu} d \sum_{\deg(\pi)=d} \eta(\pi)^{\nu/d}
\end{aligned}$$

by the definition of η. Now, multiplying by T^ν and summing over $\nu \geqslant 1$ leads to

$$T \frac{Z'}{Z}(g, \psi; T) = \sum_{\pi} \deg(\pi) \frac{\eta(\pi) T^{\deg(\pi)}}{1 - \eta(\pi) T^{\deg(\pi)}}$$

which immediately implies the result. □

Formally, we can now write down the expansion of $Z(g, \psi; T)$ as a formal power series in T, from which we hope to deduce its polynomial nature: we extend the map η by multiplicativity to all monic polynomials $f \in \mathbf{F}_q[X]$ by

$$\eta(\pi^r) = \eta(\pi)^r, \qquad \eta(f \cdot g) = \eta(f)\eta(g),$$

and we then have

$$Z(g, \psi; T) = \sum_{f} \eta(f) T^{\deg(f)} = \sum_{n \geqslant 0} c_n(\eta) T^n$$

where the sum runs over monic polynomials $f \in \mathbf{F}_q[X]$, and

$$c_n(\eta) = \sum_{\deg(f)=n} \eta(f).$$

The statement we need to prove to get Proposition 5.9 is that

$$c_n(\eta) = 0$$

for all $n \geqslant d$. Although this could be done more quickly than the way we will prove it, our argument is constructed so that it is a specialization of much more general constructions which are very important when looking at the further development of the theory.

In the previous chapter, this result was proved for a Dirichlet character by using essentially its *periodicity* modulo some polynomial (see Proposition 4.5). To obtain a similar property, we start by interpreting η as the restriction to (monic) polynomials of a character of a certain group related to the group $\mathbf{F}_q(X)^\times$ of non-zero rational functions.

Precisely, we define

$$P(\mathbf{F}_q) = \mathbf{F}_q(X)^\times / \mathbf{F}_q^\times,$$

where $\mathbf{F}_q^\times$ is seen as the subgroup of constant functions. Any element in $P(\mathbf{F}_q)$ can be represented in a unique manner as a quotient $f = f_1/f_2$ with f_i non-zero monic polynomials in $\mathbf{F}_q[X]$ and f_1 coprime with f_2. (Indeed, for unicity, a relation $f_1/f_2 = g_1/g_2$ implies $f_1g_2 = f_2g_1$, hence f_2 divides g_2, g_2 divides f_2, so that $f_2 = \alpha g_2$ for some $\alpha \in \mathbf{F}_q^\times$, and $\alpha = 1$ since they are monic; similarly for $f_1 = g_1$). In particular, as an abstract group, we can write

$$P(\mathbf{F}_q) \simeq \bigoplus_{\pi} \pi^{\mathbf{Z}}$$

where π runs as usual over monic irreducible polynomials; here the direct sum notation still corresponds to a *product* of powers of irreducible polynomials, but it expresses the fact that there are only finitely many non-trivial factors in the decomposition of some $f \in P(\mathbf{F}_q)$.

Because of this relation, the group $P(\mathbf{F}_q)$ has many characters. Indeed, for any family $\vartheta = (e^{i\theta_\pi})_\pi$, indexed by irreducible polynomials, of elements of the unit circle $\mathbf{S}^1 \subset \mathbf{C}^\times$, we can define a character

$$\eta_\vartheta \;:\; P(\mathbf{F}_q) \to \mathbf{C}^\times$$

by mapping any generator π to $e^{i\theta_\pi}$; hence we have

$$\eta_\vartheta\Bigl(\prod_\pi \pi^{m(\pi)}\Bigr) = \prod_\pi e^{im(\pi)\theta_\pi}.$$

Because there are no relations between the generators, these are all well-defined. We then note that the map η defined above from the additive exponential sums $S_\nu(g,\psi)$ is the restriction to monic polynomials of such a character, with

$$\vartheta = \Bigl(\psi\Bigl(\mathrm{Tr}_{\mathbf{F}_{q^{\deg(\pi)}}/\mathbf{F}_q}(g(\alpha_\pi))\Bigr)\Bigr)_\pi.$$

We can associate a formal Euler product

$$L(\eta_\vartheta, T) = \prod_\pi (1 - \eta_\vartheta(\pi)T^{\deg(\pi)})^{-1} = \prod_\pi (1 - e^{i\theta_\pi}T^{\deg(\pi)})^{-1}$$

to any character of $P(\mathbf{F}_q)$. *It is certainly not the case that all such Euler products can be expressed as polynomials*, even if the character is assumed non-trivial. This property requires a (rare) special feature which expresses some kind of "global" correlation between the various "local" components $e^{i\theta_\pi}$, similar (but weaker) to the periodicity of Dirichlet characters. This property is reflected by the kernel of η_ϑ being of a certain shape.

In our case, this periodicity is given by the following lemma. Before stating it, we note that we will extend the degree map from $\mathbf{F}_q[X]$ to $P(\mathbf{F}_q)$ by additivity (i.e., $\deg(f_1/f_2) = \deg(f_1) - \deg(f_2)$ for $f_i \in \mathbf{F}_q[X]$), and that we will denote by $\mathrm{ord}_0(f)$ the order of

$f \in \mathbf{F}_q(X)^\times$ at 0 (which is the multiplicity of 0 as a zero of f if f does not have a pole at 0, and is the opposite of the order of the pole otherwise), and by $\mathrm{ord}_\infty(f)$ the opposite of $\deg(f)$. The following properties are then easy to check, if not already known to the reader:

$$\mathrm{ord}_\infty f(X) = \mathrm{ord}_0 f(X^{-1}), \qquad \mathrm{ord}_\infty(f+g) \geqslant \min(\mathrm{ord}_\infty(f), \mathrm{ord}_\infty(g))$$
$$\mathrm{ord}_0(f_1 f_2) = \mathrm{ord}_0(f_1) + \mathrm{ord}_0(f_2).$$

Lemma 5.11 (Kernel of characters). *Let $\mathbf{F}_q$ be a finite field, ψ a non-trivial character of $\mathbf{F}_q$, $g \in \mathbf{F}_q[X]$ a non-constant polynomial in $\mathbf{F}_q[X]$ of degree d, with $g(0) = 0$. Let η be the character of $P(\mathbf{F}_q)$ defined by*

$$\eta(\pi) = \psi\Big(\mathrm{Tr}_{\mathbf{F}_{q^{\deg(\pi)}}/\mathbf{F}_q}(g(\alpha_\pi))\Big)$$

for any monic irreducible polynomial, where α_π is any root of π. Then we have:

(1) *The kernel of η contains the element $X \in P(\mathbf{F}_q)$.*

(2) *The kernel of η contains the subgroup $P_\infty(d+1)$ defined by*

$$P_\infty(d+1) = \{f \in P(\mathbf{F}_q) \mid \mathrm{ord}_\infty(X^{-\deg(f)}f(X) - 1) \geqslant d+1\}.$$

Equivalently, $\mathrm{Ker}\,\eta$ *contains the subgroup $P_0(d+1)$ defined by*

$$P_0(d+1) = \{f \in P(\mathbf{F}_q) \mid \mathrm{ord}_0(X^{\deg(f)}f(X^{-1}) - 1) \geqslant d+1\}.$$

(3) *If $(d, q) = 1$, then η is a non-trivial character of $P(\mathbf{F}_q)$.*

Proof. (1) Since X is irreducible, we have $\eta(X) = \psi(g(0)) = \psi(0) = 1$ by definition and by the assumption $g(0) = 0$.

(2) We first check that $P_\infty(k)$ is indeed a subgroup of $P(\mathbf{F}_q)$ for any integer $k \geqslant 0$: this follows from the formulas

$$\begin{aligned}\mathrm{ord}_\infty(X^{-\deg(f)}f - 1) &= \mathrm{ord}_\infty(X^{-\deg(f)}f) + \mathrm{ord}_\infty(1 - X^{\deg(f)}f^{-1})\\ &= \mathrm{ord}_\infty(X^{\deg(f)}f^{-1} - 1),\end{aligned}$$

and

$$\begin{aligned}\mathrm{ord}_\infty(X^{-\deg(f_1f_2)}f_1f_2 - 1) &= \mathrm{ord}_\infty\Big((X^{-\deg(f_1)}f_1 - 1)\\ &\qquad + X^{-\deg(f_1)}f_1(X^{-\deg(f_2)}f_2 - 1)\Big)\\ &= \mathrm{ord}_\infty(h_1 + h_2)\end{aligned}$$

where $\mathrm{ord}_\infty(h_1) \geqslant k$ if $f_1 \in P_\infty(k)$, while also

$$\begin{aligned}\mathrm{ord}_\infty(h_2) &= \mathrm{ord}_\infty(X^{-\deg(f_1)}f_1 \times (X^{-\deg(f_2)}f_2 - 1)))\\ &= \mathrm{ord}_\infty(X^{-\deg(f_2)}f_2 - 1) \geqslant k\end{aligned}$$

if $f_2 \in P_\infty(k)$.

Now to come back to the main point, let $f \in P(\mathbf{F}_q)$ be written $f = f_1/f_2$ with $f_i \in \mathbf{F}_q[X]$ coprime monic polynomials. Then we have

$$X^{-\deg(f)}f(X) = \frac{X^{-\deg(f_1)}f_1(X)}{X^{-\deg(f_2)}f_2(X)}$$

and hence

$$X^{-\deg(f)}f(X) - 1 = \frac{X^{-\deg(f_1)}f_1(X) - X^{-\deg(f_2)}f_2(X)}{X^{-\deg(f_2)}f_2(X)}.$$

Let $n = \deg(f_1)$, $m = \deg(f_2)$, and write

$$f_1 = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0,$$
$$f_2 = X^m + b_{m-1}X^{m-1} + \cdots + b_1X + b_0,$$

so that

$$X^{-n}f_1 = 1 + a_{n-1}X^{-1} + \cdots + a_0X^{-n},$$
$$X^{-m}f_2 = 1 + b_{m-1}X^{-1} + \cdots + b_0X^{-m}.$$

Note that $\mathrm{ord}_\infty(X^{-m}f_2) = 0$ and therefore $f \in P_\infty(d+1)$ if and only if

$$\mathrm{ord}_\infty(X^{-n}f_1 - X^{-m}f_2) \geqslant d+1,$$

and by definition this happens if and only if the coefficients of 1, X^{-1}, X^{-2}, …, X^{-d} in the difference $X^{-n}f_1 - X^{-m}f_2$ are all zero, namely when

$$a_{n-1} = b_{m-1}, \quad \ldots, \quad a_{n-d} = b_{m-d} = 0, \tag{5.4}$$

with the convention $a_i = b_i = 0$ if $i < 0$.

With this in mind, we now observe that for a monic polynomial $h \in \mathbf{F}_q[X]$, we can express concisely $\eta(h)$ as follows: we have

$$\eta(h) = \psi\Bigl(\sum_{h(\alpha)=0} g(\alpha)\Bigr),$$

where α runs over all the roots of h, counted with multiplicity. Indeed, this is because this expression is itself multiplicative with respect to h, and for $h = \pi$ irreducible, we have

$$\sum_{\pi(\alpha)=0} g(\alpha) = \mathrm{Tr}_{\mathbf{F}_{q^{\deg(\pi)}}/\mathbf{F}_q}(g(\alpha_\pi)).$$

Now consider the map

$$F \,:\, h \mapsto \sum_{h(\alpha)=0} g(\alpha).$$

Since $g \in \mathbf{F}_q[X]$ is a fixed polynomial, this is obviously a symmetric function of the roots of h, in fact a linear combination of the sums of k-th powers of the roots, for $1 \leqslant k \leqslant \deg(g)$; there is no constant term because $g(0) = 0$. Consequently, by the theory of symmetric functions, $F(h)$ is also a polynomial in the $\deg(g)$ first elementary symmetric functions of the roots, i.e., the functions

$$\sum_{\alpha} \alpha, \qquad \sum_{\alpha\neq\beta} \alpha\beta, \qquad \ldots,$$

which (up to signs) are, in terms of the coefficients of

$$h = \sum_{i\geqslant 0} c_i(h)X^i,$$

simply the functions $e_i \,:\, h \mapsto c_{\deg(h)-i}(h)$, for $1 \leqslant i \leqslant \deg(g)$. Hence the condition (5.4) is in fact equivalent with $F(f_1) = F(f_2)$, in which case

$$\eta(f) = \eta(f_1)\eta(f_2)^{-1} = \psi(F(f_1) - F(f_2)) = 1,$$

as claimed.

(3) We try construct a polynomial f where $\eta(f) \neq 1$ by taking

$$f = X^d - \beta$$

for some $\beta \in \mathbf{F}_q$. In that case, the roots are given by $\alpha = \xi\alpha_0$ where α_0 is any root of $\alpha_0^d = \beta$ and ξ runs over the d-th roots of unity in $\bar{\mathbf{F}}_q$. Thus, we have

$$\sum_{f(\alpha)=0} \alpha^j = \alpha_0^j \sum_{\xi^d=1} \xi^j$$

and this is zero for $1 \leqslant j \leqslant d-1$. Therefore, expanding g is powers of X, we find

$$\sum_{f(\alpha)=0} g(\alpha) = \gamma \sum_{\xi^d=1} \alpha_0^d \xi^d = d\beta\gamma,$$

where $\gamma \neq 0$ is the leading coefficient of $g \in \mathbf{F}_q[X]$. Such polynomials therefore satisfy

$$\eta(f) = \psi(d\beta\gamma).$$

Now, since ψ is non-trivial, $\gamma \neq 0$ and d is assumed to be coprime with q, we can find β such that $d\beta\gamma$ is an element of $\mathbf{F}_q$ with $\psi(d\beta\gamma) \neq 1$, and this shows that η is non-trivial. □

Example 5.12. It is quite easy to see how the argument above translates to explicit expressions for the character associated to concrete polynomials. Consider for instance $g = X^3 + aX$ where $a \in \mathbf{F}_q$. For a monic polynomial $h \in \mathbf{F}_q[X]$, which we write as

$$h = X^n + a_1 X^{n-1} + a_2 X^{n-2} + a_3 X^{n-3} + \cdots + a_{n-1} X + a_n = \prod_{h(\alpha)=0} (X - \alpha),$$

we have the identities

$$\sum_{\alpha} \alpha = -a_1, \qquad \sum_{\alpha} \alpha^3 = -a_1^3 - 3a_3 + 3a_1 a_2,$$

and hence

$$\eta(h) = \psi(-a_1^3 - 3a_3 + 3a_1 a_2 - a a_1).$$

Of course, although this expression is very concrete, it is not clear from a simple look that it is multiplicative with respect to η!

The situation is now quite favorable because we can see η as a character of the quotient group modulo $P_\infty(d+1)$:

$$\eta : P(\mathbf{F}_q) \to P(\mathbf{F}_q)/P_\infty(d+1) \to \mathbf{C}^\times.$$

As one can expect, this quotient is finite. More precisely, analyzing it will quickly lead to:

Proof of Proposition 5.9. We denote by $C(d+1)$ the quotient group

$$C(d+1) = P(\mathbf{F}_q)/P_\infty(d+1),$$

and identify η with a character of this group. First of all, we claim that this is a finite group, and that any class $c \in C(d+1)$ has a *unique* representative which is a monic polynomial of degree d. Indeed, let c be the class of f_1/f_2 with f_i coprime monic polynomials, given by

$$f_i = X^{n_i} + a_{i,1} X^{n_i - 1} + \cdots + a_{i,n_i-1} X + a_{i,n_i},$$

and let

$$h = X^d + b_1 X^{d-1} + \cdots + b_{d-1} X + b_d$$

be a polynomial of degree d. Then the class of h is equal to c if and only if

$$\mathrm{ord}_\infty\Big(\frac{hf_2}{f_1} - 1\Big) \geqslant d+1.$$

As in the proof of Lemma 5.11, this translates exactly to equations expressing that the $d+1$ coefficients of topmost degree of hf_2 and f_1 coincide, namely:

$$b_1 + a_{2,1} = a_{1,1}, \quad b_2 + a_{2,1}b_1 + a_{2,2} = a_{1,2}, \quad \ldots \\ b_d + a_{2,1}b_{d-1} + \cdots + a_{2,d-1}b_1 + a_{2,d} = a_{1,d}$$

This linear system of equations (with unknowns the b_i's) has a solution, which shows the existence of a representative as claimed. For unicity, note that two different monic polynomials h_1, h_2 of degree d can not have $h_1/h_2 \in P_\infty(d+1)$, by the same argument based on the proof of Lemma 5.11. (In particular, we see that $C(d+1)$ is a group of order q^d.)

Now let $n \geqslant d$ be given, and let

$$c_n(\eta) = \sum_{\deg(f)=n} \eta(f)$$

be the coefficient of T^n in $Z(g,\psi;T)$ (the sum ranging, of course, over monic polynomials in $\mathbf{F}_q[X]$). We have

$$c_n(\eta) = \sum_{c\in C(d+1)} \eta(c) \sum_{\substack{\deg(f)=n\\ f\equiv c\,(\mathrm{mod}\,P_\infty(d+1))}} 1.$$

If we consider the unique representative h of c constructed above (of degree d, with coefficients b_i), the congruence $f \equiv h\,(\mathrm{mod}\,P_\infty(d+1))$ means that

$$f = X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n,$$

with

$$a_1 = b_1, \quad a_2 = b_2, \quad \cdots, \quad a_d = b_d.$$

Thus these coefficients are fixed (and can be fixed in this manner because $n \geqslant d$), while all the other coefficients (namely a_{d+1}, ..., a_n) can be freely chosen in $\mathbf{F}_q$. This means that the inner sum is equal to q^{n-d} for all c, and therefore

$$c_n(\eta) = q^{n-d} \sum_{c\in C(d+1)} \eta(c) = 0$$

since η is, for $(d,q)=1$, a non-trivial character of the finite quotient group $C(d+1)$ by Lemma 5.11, (3). □

Example 5.13. Continuing the previous example, with $g = X^3 + aX$ and

$$\eta(X^n + a_1X^{n-1} + \cdots + a_{n-1}X + a_n) = \psi(-a_1^3 - 3a_3 + 3a_1a_2 - aa_1),$$

where $a \in \mathbf{F}_q$ is fixed, the computation boils down to

$$\sum \cdots \sum_{(a_1,\ldots,a_n)\in\mathbf{F}_q} \psi(-a_1^3 - 3a_3 + 3a_1a_2 - aa_1) = 0$$

if $n \geqslant 3$, which is of course quite clear if $3 \nmid q$, because the free summation over a_3 can be extracted and gives 0:

$$\sum \cdots \sum_{(a_1,\ldots,a_n)\in\mathbf{F}_q} \psi(-a_1^3 - 3a_3 + 3a_1a_2 - aa_1) = \\ q^{n-3} \sum_{a_1,a_2\in\mathbf{F}_q} \psi(-a_1^3 + 3a_1a_2 - aa_1) \sum_{a_3} \psi(-3a_3) = 0.$$

Remark 5.14. (1) More intrinsically, the group $P(\mathbf{F}_q)$ should be seen as the group of *fractional ideals* in the field $\mathbf{F}_q(X)$. Because $\mathbf{F}_q[X]$ is a principal ideal domain, such ideals are all principal, and thus a fractional ideal corresponds to a non-zero element $f \in \mathbf{F}_q(X)^\times$, modulo the group of units $\mathbf{F}_q^\times$ (multiplying by units corresponds to the different possible generators of a fractional ideal).

(2) We can also see Dirichlet characters as corresponding to certain characters of certain *subgroups* of $P(\mathbf{F}_q)$. This would allow (as in [18]) a more direct approach of both types of sums simultaneously (including even, quite easily, mixed character sums, that we will not discuss explicitly in this first part).

(3) As the two previous remarks suggest, an even better understanding of the situation would involve the detailed study of Hecke Grössencharakters of all "global function fields"; this would involve defining them as suitable characters of idèle groups of finite extensions of $\mathbf{F}_q(X)$.

5.4. Kloosterman sums in odd characteristic

Theorem 3.2 does not cover the important special case of estimating Kloosterman sums (i.e., it does not imply the Weil bound, Theorem 10 of the introduction) because the latter involves the rational function $aX + bX^{-1}$ in the additive character.

CHAPTER 6

Heilbronn sums

6.1. Introduction

In the two previous chapters, we have proved important special cases of the Riemann Hypothesis over finite fields for one-variable exponential sums. The results are best possible in some sense (in particular, once the degree of the L-function, as a polynomial, is known, as well as the fact that the inverse roots are Weil numbers of weight 1, one can not improve on the estimate for the companion sums over extension fields $\mathbf{F}_{q^\nu}$ as $\nu \to +\infty$). However, in applications, one may have to deal with exponential sums for which the Riemann Hypothesis does not lead to any non-trivial result. A typical example is given by the sums

$$G_k(a;p) = \sum_{x\in\mathbf{F}_p} e\Bigl(\frac{ax^k}{p}\Bigr), \qquad a \in \mathbf{F}_p^\times,$$

if k is not considered as fixed, but is allowed to vary with p. Indeed, if $k \mid p-1$, detecting k-th powers in $\mathbf{F}_q$ using multiplicative characters, one gets

$$\begin{aligned}\sum_{x\in\mathbf{F}_p} e\Bigl(\frac{ax^k}{p}\Bigr) &= \sum_{y\in\mathbf{F}_p} e\Bigl(\frac{ay}{p}\Bigr)|\{x\in\mathbf{F}_p \mid x^k = y\}| \\ &= \sum_{y\in\mathbf{F}_q} e\Bigl(\frac{ay}{p}\Bigr)\sum_{\chi^k=1}\chi(y) \\ &= \sum_{\chi^k=1}\sum_{y\in\mathbf{F}_q}\chi(y)e\Bigl(\frac{ay}{p}\Bigr) \\ &= \sum_{\substack{\chi^k=1\\ \chi\neq 1}}\tau(\chi,\psi_a)\end{aligned}$$

where $\psi_a(x) = e(ax/p)$ is a non-trivial additive character of $\mathbf{F}_p$; we have used the fact that $\tau(1,\psi_a) = 0$. This expression is a sum of p-Weil numbers of weight 1; however, the bound it leads to is

$$\Bigl|\sum_{x\in\mathbf{F}_p} e\Bigl(\frac{ax^k}{p}\Bigr)\Bigr| \leqslant (k-1)\sqrt{p}$$

(which is also what the Riemann Hypothesis gives here), and if $k > \sqrt{p}$, this is no better, or even worse, than the trivial bound

$$|G_k(a;p)| = \Bigl|\sum_{x\in\mathbf{F}_p} e\Bigl(\frac{ax^k}{p}\Bigr)\Bigr| \leqslant p.$$

Even independently of potential applications (which certainly exist), it is natural to take this as a challenge to our understanding of exponential sums. Is this trivial bound really the best that can be done? If not, how could one improve it?

In this chapter, we will consider another type of sum, which is however closely related, and for which the same type of questions arise.

Definition 6.1 (Heilbronn sum). Let p be a prime number and let $a \in \mathbf{F}_p^\times$. The *Heilbronn sum* $H(a;p)$ is defined by

$$H(a;p) = \sum_{x \in \mathbf{F}_p^\times} e\Bigl(\frac{ax^p}{p^2}\Bigr). \tag{6.1}$$

Note that this is not quite an exponential sum of the type we have considered up to now, since the denominator p^2 is not a typo (if it were replaced by p, since $x^p \equiv x \,(\mathrm{mod}\, p)$, the sum would be -1). So a first remark is in order to justify the definition: the point is that if $n \in \mathbf{Z}$ is any lift of $x \in \mathbf{F}_p^\times$, we *define*

$$e\Bigl(\frac{ax^p}{p^2}\Bigr) = e\Bigl(\frac{an^p}{p^2}\Bigr),$$

and this is well-defined because if $m = n + \alpha p$ is any other integer reducing to x modulo p, we have

$$m^p = (n+\alpha p)^p = n^p + p \cdot (\alpha p) n^{p-1} + \binom{p}{2}(\alpha p)^2 n^{p-2} + \cdots$$

and this is congruent to n^p modulo p^2, so that

$$e\Bigl(\frac{am^p}{p^2}\Bigr) = e\Bigl(\frac{an^p}{p^2}\Bigr),$$

which shows that the sum is well-defined.

The only obvious bound for this sum is

$$|H(a;p)| \leqslant p - 1.$$

Can one do better? This question was raised by Heilbronn in the 1940's, and only recently was there some progress, due to Heath-Brown [8]. We will prove his result:

Theorem 6.2 (Heath-Brown). *We have*

$$H(a;p) \ll p^{11/12}$$

for all primes p and $a \in \mathbf{F}_p^\times$, where the implied constant is absolute. In fact, more precisely, we have

$$|H(a;p)| \leqslant 4p^{11/12}. \tag{6.2}$$

Remark 6.3. This is not the best result currently known: by elaborating on Heath-Brown's ideas, Heath-Brown and Konyagin [9] have shown

$$H(a;p) \ll p^{7/8},$$

and indeed they have also proved that

$$G_k(a;p) \ll \min((kp)^{5/8}, k^{3/8}p^{3/4})$$

using similar techniques, which is a non-trivial estimate as long as $k \ll p^{2/3}$, whereas the Riemann Hypothesis bound is only non-trivial for $k \ll p^{1/2}$.

The methods involved are elementary and quite closely related to Stepanov's method, in its original form. The strategy is to reduce first to point-counting on some curve over $\mathbf{F}_p$; the fact that the sum bears some similarity to the sums $G_k(a;p)$ appears clearly at this point, since the curve involved has large degree depending on p, so that the point counting results deriving from the Riemann Hypothesis would only lead to a trivial bound, once more. Then a variant of Stepanov's method is shown to be applicable to the point-counting problem.

6.2. Reduction to point counting

In this section, we show how to go from the problem of estimating Heilbronn sums to that of counting solutions to (large degree) polynomial equations over $\mathbf{F}_p$. This is given by the following proposition:

Proposition 6.4. *Let p be a prime number. Define the polynomial*

$$L_p = X + \frac{X^2}{2} + \cdots + \frac{X^{p-1}}{p-1} \in \mathbf{F}_p[X], \tag{6.3}$$

and let

$$\mathcal{N}_r(\mathbf{F}_p) = \{x \in \mathbf{F}_p - \{0,1\} \mid L(x) = r\}.$$

for $r \in \mathbf{F}_p$.

Then for any $a \in \mathbf{F}_p^\times$, we have

$$|H(a;p)| \leqslant (p-1)^{1/2} + \mathcal{N}(\mathbf{F}_p)^{1/4} p^{3/4}$$

where

$$\mathcal{N}(\mathbf{F}_p) = \max_{r \in \mathbf{F}_p} |\mathcal{N}_r(\mathbf{F}_p)|,$$

and the implied constant is absolute.

Remark 6.5. It is important to note that although this estimate does not lead to any direct upper-bound for the Heilbronn sums by itself, it remains "neutral": if we input the obvious bound $|\mathcal{N}_r(\mathbf{F}_p)| \leqslant p$ for all r, the right-hand side is of size p, which is the trivial bound for Heilbronn sums. In other words this proposition has not led to any direct loss for the estimate. Thus one may get a result from any non-trivial understanding of the order of $\mathcal{N}_r(\mathbf{F}_p)$. This is the point-counting problem that will be dealt with in the next section.

The second remark is that the polynomial involved, L_p, is a truncated logarithm: we have

$$\log\Bigl(\frac{1}{1-X}\Bigr) = X + \frac{X^2}{2} + \cdots + \frac{X^n}{n} + \cdots$$

in the formal power-series ring $\mathbf{Q}[[X]]$, and L_p is obtained by eliminating the terms of degree $\geqslant p$ – since X^p/p does not make sense in $\mathbf{F}_p[X]$, this is necessary. We will see that, in quite a few places, this "transcendental" origin of the polynomial plays a crucial role.

Proof. The basic idea is quite similar to well-known techniques in the study of exponential sums: one writes down $|H(a;p)|^2$ and tries to rearrange the double sum that arises in terms of new variables.

We have

$$|H(a;p)|^2 = \sum_{x,y \in \mathbf{F}_p^\times} e\Bigl(\frac{a(x^p - y^p)}{p^2}\Bigr),$$

and we start by isolating the “diagonal” contribution where $x = y$, then for the remainder we put

$$y = x - t$$

where $t \in \mathbf{F}_p - \{0, x\}$: we get

$$|H(a; p)|^2 = p - 1 + \sum_{x\in\mathbf{F}_p^\times} \sum_{t\in\mathbf{F}-\{0,x\}} e\Bigl(\frac{a(x^p - (x-t)^p)}{p^2}\Bigr).$$

Then it seems natural to factor x^p in the exponential; thus we write $t = ux$ with $u \in \mathbf{F}_p - \{0, 1\}$, and get

$$|H(a; p)|^2 = p - 1 + \sum_{x\in\mathbf{F}_p^\times} \sum_{u\in\mathbf{F}-\{0,1\}} e\Bigl(\frac{a(x^p(1 - (1-u)^p)}{p^2}\Bigr).$$

We now expand the term $1 - (1 - u)^p$ and reduce it modulo p^2; this is similar to the Taylor expansions often performed in the Weyl or van der Corput methods. We may of course assume that $p \neq 2$, and thus we have

$$\begin{aligned} 1 - (1-u)^p &= -\sum_{j=1}^{p} (-1)^j \binom{p}{j} u^j \\ &= (-1)^{p+1} u^p - \sum_{j=1}^{p-1} (-1)^j \binom{p}{j} u^j \\ &= u^p - \sum_{j=1}^{p-1} (-1)^j \binom{p}{j} u^j. \end{aligned}$$

However, for $1 \leqslant j \leqslant p - 1$, we can write

$$\begin{aligned} \binom{p}{j} &= \frac{p(p-1)\cdots(p-j+1)}{1\cdot 2\cdots(j-1)\cdot j} \\ &= p\cdot\frac{p-1}{1}\cdot\frac{p-2}{2}\cdots\frac{p-(j-1)}{j-1}\cdot\frac{1}{j} \\ &\equiv (-1)^{j-1}\frac{p}{j} \pmod{p^2}, \end{aligned}$$

and consequently we get

$$1 - (1-u)^p \equiv u^p + p\sum_{j=1}^{p-1} \frac{u^j}{j} = u^p + pL_p(u) \pmod{p^2}.$$

This means that

$$|H(a; p)|^2 = p - 1 + \sum_{x\in\mathbf{F}_p^\times} \sum_{u\in\mathbf{F}-\{0,1\}} e\Bigl(\frac{ax^p(u^p + pL_p(u))}{p^2}\Bigr).$$

Now, note that if $u \neq 0$ in $\mathbf{F}_p$, we have $u^{p-1} \equiv 1 \pmod p$ and hence

$$\begin{aligned} L_p(u) = \Bigl(u + \frac{u^2}{2} + \cdots + \frac{u^{p-1}}{p-1}\Bigr) &= u^p\Bigl(u^{1-p} + \frac{u^{2-p}}{2} + \cdots + \frac{u^{-1}}{p-1}\Bigr) \\ &= u^p\Bigl(-\frac{u^{1-p}}{p-1} - \frac{u^{p-2}}{p-2} - \cdots - u^{-1}\Bigr) \\ &= -u^pL_p(u^{-1}) \end{aligned}$$

in $\mathbf{F}_p$ (which could be compared with the equation $\log(x^{-1}) = -\log x$). This leads to the new expression

$$|H(a;p)|^2 = p - 1 + \sum_{x\in\mathbf{F}_p^\times} \sum_{u\in\mathbf{F}-\{0,1\}} e\Bigl(\frac{a(ux)^p(1-pL_p(u^{-1}))}{p^2}\Bigr).$$

We are now naturally led to make the change of variable $(v = u^{-1}, y = ux)$ with $v \in \mathbf{F}_p - \{0,1\}$ and $y \in \mathbf{F}^\times$, and we obtain

$$\begin{aligned} |H(a;p)|^2 &= p - 1 + \sum_{y\in\mathbf{F}_p^\times} \sum_{v\in\mathbf{F}-\{0,1\}} e\Bigl(\frac{a(1-pL_p(v))y^p}{p^2}\Bigr) \\ &= p - 1 + \sum_{r\in\mathbf{F}_p} \sum_{\substack{v\in\mathbf{F}_p-\{0,1\} \\ L_p(v)=r}} \sum_{y\in\mathbf{F}^\times} e\Bigl(\frac{a(1-pr)y^p}{p^2}\Bigr) \\ &= p - 1 + \sum_{r\in\mathbf{F}_p} H(a(1-pr);p)|\mathcal{N}_r(\mathbf{F}_p)|. \end{aligned}$$

This is the basic relation we need. Now, to obtain the proposition, we must separate the Heilbronn sums $H(a(1-pr);p)$ from the point-counting values $\mathcal{N}_r(\mathbf{F}_p)$. For this, we first note the obvious relation

$$\sum_{r\in\mathbf{F}_p} |\mathcal{N}_r(\mathbf{F}_p)| = |\mathbf{F}_p - \{0,1\}| = p - 2, \tag{6.4}$$

and the mean-square estimate

$$\begin{aligned} \sum_{r\in\mathbf{F}_p} |H(a(1-pr);p)|^2 &= \sum_{r\in\mathbf{F}_p} \sum_{x,y\in\mathbf{F}_p^\times} e\Bigl(\frac{a(1-pr)(x^p-y^p)}{p^2}\Bigr) \\ &= \sum_{x,y\in\mathbf{F}_p^\times} e\Bigl(\frac{a(x^p-y^p)}{p^2}\Bigr) \sum_{r\in\mathbf{F}_p} e\Bigl(\frac{-ar(x^p-y^p)}{p}\Bigr) \\ &= p(p-1) \end{aligned}$$

by orthogonality of characters and the fact that $x^p = y^p$ implies $x = y$ in $\mathbf{F}_p$.

Then we apply Cauchy's inequality to the second term of the expression we obtained for $|H(a;p)|^2$, to derive from these two facts the upper bound

$$\begin{aligned} \Bigl|\sum_{r\in\mathbf{F}_p} H(a(1-pr);p)|\mathcal{N}_r(\mathbf{F}_p)|\Bigr|^2 &\leqslant \sum_{r\in\mathbf{F}_p} |\mathcal{N}_r(\mathbf{F}_p)|^2 \sum_r |H(a(1-pr);p)|^2 \\ &\leqslant p(p-1)(p-2) \max_{r\in\mathbf{F}_p} |\mathcal{N}_r(\mathbf{F}_p)|. \end{aligned}$$

This leads immediately to the proposition since $\sqrt{a^2+b^2} \leqslant a + b$ for any a, $b \geqslant 0$. □

6.3. Point counting for Heilbronn sums

From the previous section, we see that Theorem 6.2 will follow from the next result. The exponent of p is obtained by

$$\frac{1}{4}\cdot\frac{2}{3}+\frac{3}{4}=\frac{11}{12},$$

and the more precise bound (6.2) comes from

$$|H(a;p)|\leqslant\sqrt{p}+44^{1/4}p^{11/12}\leqslant\sqrt{p}+3p^{11/12}\leqslant 4p^{11/12}$$

for all $p\geqslant 2$ and $a\in\mathbf{F}_p^{\times}$.

Theorem 6.6 (Mit'kin; Heath-Brown). *Let p be a prime number, and let*

$$L_p=X+\frac{X^2}{2}+\cdots+\frac{X^{p-1}}{p-1}\in\mathbf{F}_p[X],$$

and let

$$\mathcal{N}_r(\mathbf{F}_p)=\{x\in\mathbf{F}_p-\{0,1\}\mid L_p(x)=r\}.$$

for $r\in\mathbf{F}_p$. Then we have

$$|\mathcal{N}_r(\mathbf{F}_p)|\ll p^{2/3}$$

where the implied constant is absolute. In fact, we have

$$|\mathcal{N}_r(\mathbf{F}_p)|\leqslant 44p^{2/3}. \tag{6.5}$$

Although this result was proved also by Heath-Brown in his paper [8], it turned out that this point-counting result (but not its application to Heilbronn sums) had been proved already by D.A. Mit'kin [17] just a few years before.

The proof is elementary, but delicate. To understand a bit why it is not unreasonable to expect this to be true (or at least something similar), note that we have already seen in (6.4) that the average of $|\mathcal{N}_r(\mathbf{F}_p)|$ is bounded (by 1). However, we are interested in a bound on the maximal size, and in order for this to be also reasonably small, we may look at the mean square of $|\mathcal{N}_r(\mathbf{F}_p)|$ (in fact, it is from this mean-square that we arrived at the maximum). Expanding the square, we find that

$$\begin{aligned}\sum_{r\in\mathbf{F}_p}|\mathcal{N}_r(\mathbf{F}_p)|^2&=\sum_{x,y\in\mathbf{F}_p-\{0,1\}}\ \sum_{\substack{r\in\mathbf{F}_p\\ L_p(x)=L_p(y)=r}}1\\&=|\{(x,y)\in(\mathbf{F}_p-\{0,1\})^2\mid L_p(x)=L_p(y)\}|.\end{aligned}$$

This is therefore the number of points, with coordinates in $\mathbf{F}_p-\{0,1\}$, of the curve defined by the equation

$$L_p(X)-L_p(Y)=0$$

in the plane. There are of course $p-2$ solutions corresponding to diagonal pairs (x,x), but for the remaining ones, the fact that the degree of the polynomial $L_p(X)-L_p(Y)$ is large (it is $p-1$) means that even if we could apply directly the Riemann Hypothesis for the number of solutions to this equation, we would only get a trivial estimate of size roughly $p^{5/2}$ (the genus of the curve could be as large as p^2). So the problem is quite similar to what we described at the beginning of this chapter with the estimates of $G_k(a;p)$.

Heath-Brown's adaptation of the Stepanov method to deal with this problem is based on the construction of an auxiliary polynomial $F\in\mathbf{F}_p[A,B,C]$, in three variables A, B and C, with the following properties:

(i) its degree with respect to each variable will not be "too large";

(ii) the associated specialized polynomial

$$G = F(X, X^p, L_p(X)) \in \mathbf{F}_p[X]$$

is non-zero, and is such that G vanishes to a fairly larger order $\geqslant m$ for all $x \in \mathcal{N}_r(\mathbf{F}_p)$, i.e., all $x \in \mathbf{F}_p - \{0, 1\}$ such that $L_p(x) = r$ (where r is fixed and F may of course depend on r).

It follows in that case that

$$|\mathcal{N}_r(\mathbf{F}_p)| \leqslant \frac{\deg(G)}{m} \leqslant \frac{\deg_A(F) + p\deg_B(F) + (p-1)\deg_C(F)}{m}. \tag{6.6}$$

Since a bound of size p is trivial, we see already that it will be imperative that m be quite large (a positive power of p). However, we will also see that one must take $m < p$, and this allows us to use usual derivatives (instead of Hasse derivatives) to detect zeros of order $\geqslant m$:

Lemma 6.7. *Let K be a field of characteristic $p > 0$, $f \in K[X]$ a polynomial. Then, if $0 \leqslant m \leqslant p$, an element $x \in K$ is a zero of f of order $\geqslant m$ if and only if*

$$f(x) = f'(x) = \cdots = f^{(m-1)}(x) = 0.$$

If $x \neq 0, 1$, then x is a zero of f of order $\geqslant m$ if and only if

$$f(x) = \delta f(x) = \cdots = \delta^{m-1} f(x) = 0,$$

where δ is the linear map

$$\begin{cases} K[X] & \longrightarrow & K[X] \\ f & \mapsto & X(1-X)f' \end{cases}.$$

The example of $f = X^p$, where $x = 0$ is a zero of order p but $f^{(j)}(0) = 0$ for *all* $j \geqslant 0$ (because the p-th derivative is $p! = 0$), shows that this does not hold anymore for order $> p$. Then one would have to appeal to Hasse derivatives (see the Appendix to Chapter 4).

The main reason why the strategy above has a chance is the following obvious fact: if $x \in \mathcal{N}_r(\mathbf{F}_p)$, the auxiliary polynomial G evaluated at x satisfies

$$G(x) = H(x) \tag{6.7}$$

where H is the polynomial $H = F(X, X, r) \in \mathbf{F}_p[X]$. The point is that the degree of H is much smaller than that of G: we have

$$\deg(H) \leqslant \deg_A(F) + \deg_B(F).$$

This means, in particular, that it may well be that $H = 0$ (as a polynomial) even when $G \neq 0$. And if that is the case, we have at least the starting case $m = 1$ of condition (ii): the polynomial G then vanishes to order $\geqslant 1$ at all points of $\mathcal{N}_r(\mathbf{F}_p)$. (This argument is the analogue to Lemma 4.27 in Bombieri's variant of the Stepanov method.)

To deal with cases where $m \geqslant 1$, we use the following lemma to generalize (6.7) to derivatives of G. First of all, for convenience we denote by Φ the $\mathbf{F}_p$-linear specialization giving the auxiliary polynomial G:

$$\Phi \begin{cases} \mathbf{F}_p[A, B, C] & \longrightarrow & \mathbf{F}_p[X] \\ F & \mapsto & F(X, X^p, L_p(X)) \end{cases}.$$

Lemma 6.8. *Let p be a prime, $L_p \in \mathbf{F}_p[X]$ as before. For $F \in \mathbf{F}_p[A, B, C]$ and $G = \Phi(F) = F(X, X^p, L_p(X)) \in \mathbf{F}_p[X]$, we have*

$$\delta G = X(1-X)G' = \partial(F)(X, X^p, L_p(X)) = \Phi(\partial F),$$

where ∂ denotes the map

$$\begin{cases} \mathbf{F}_p[A,B,C] & \longrightarrow & \mathbf{F}_p[A,B,C] \\ F & \mapsto & A(1-A)\dfrac{\partial F}{\partial A} + (A-B)\dfrac{\partial F}{\partial C} \end{cases}.$$

Proof. It is enough to check this for a monomial $F = A^a B^b C^c$, by linearity. Even better: it is a simple fact that either $\Delta = \Phi \circ \partial$ or $\Delta = \delta \circ \Phi$, as $\mathbf{F}_p$-linear maps

$$\mathbf{F}_p[A,B,C] \to \mathbf{F}_p[X],$$

satisfy the following version of Leibniz rule:

$$\Delta(F_1F_2) = \Phi(F_1)\Delta(F_2) + \Phi(F_2)\Delta(F_1)$$

for F_1, $F_2 \in \mathbf{F}_p[A,B,C]$.

Using this rule, it follows immediately that the desired conclusion (namely, $\Phi \circ \partial = \delta \circ \Phi$) follows from its validity in the special cases $F = A$, $F = B$ and $F = C$.

We consider each of these in turn. For $F = A$, we have $G = \Phi(A) = X$ and

$$\delta G = X(1-X) = \Phi(\partial A),$$

and for $F = B$, we have $G = \Phi(B) = X^p$ and

$$\delta G = 0 = \Phi(\partial B).$$

Finally,for $F = C$, we have $G = \Phi(C) = L_p(X)$ and by definition (6.3), we get

$$\begin{aligned} X(1-X)G' &= X(1-X)\Big(1 + X + \cdots + X^{p-2}\Big) \\ &= X - X^p = \Phi(A-B) = \Phi(\partial C), \end{aligned}$$

as expected. □

Remark 6.9. The identity

$$X(1-X)L_p'(X) = X - X^p$$

should be compared with the standard formula

$$X(1-X)\Big(\log\frac{1}{1-X}\Big)' = X.$$

From this lemma, we see that for $j \geqslant 0$, we have

$$\delta^j G = (\partial^j F)(X, X^p, L_p(X)) \in \mathbf{F}_p[X],$$

while for $j \geqslant 0$ and $x \in \mathcal{N}_r(\mathbf{F}_p)$, we get

$$(\delta^j G)(x) = H_j(x)$$

with

$$H_j = (\partial^j F)(X, X, r).$$

Consequently, if $0 \leqslant m < p$ and if $F \in \mathbf{F}_p[A,B,C]$ has the property that the polynomials H_j defined in this manner satisfy

$$H_0 = \cdots = H_{m-1} = 0$$

(as polynomials in $\mathbf{F}_p[X]$), then any $x \in \mathcal{N}_r(\mathbf{F}_p)$ is a zero of $G = F(X, X^p, L_p(X))$ of order $\geqslant m$.

Now, counting unknowns and equations, we will quickly get:

Lemma 6.10 (Ensuring high order of vanishing). *Let $r \in \mathbf{F}_p$ be fixed. If $0 \leqslant m < p$ and a, b, $c \geqslant 1$ are integers such that*

$$m(a+b+m-1) < abc, \tag{6.8}$$

then there exists a non-zero polynomial $F \in \mathbf{F}_p[A, B, C]$ such that

$$\deg_A(F) < a, \quad \deg_B(F) < b, \quad \deg_C(F) < c,$$

and

$$F(X, X, r) = (\partial F)(X, X, r) = \cdots = (\partial^{m-1} F)(X, X, r) = 0,$$

and in particular such that each $x \in \mathcal{N}_r(\mathbf{F}_p)$ is a zero of

$$G = F(X, X^p, L_p(X))$$

of order $\geqslant m$.

This step is analogue to the use of the Riemann-Roch theorem to construct a function f satisfying the relation (4.30) in Lemma 4.27. Note that this does not yet give the existence of the desired auxiliary polynomials G: we must afterwards still ensure that one can find a polynomial F as above such that $G \neq 0 \in \mathbf{F}_p[X]$, and whether this is possible is by no means clear! (In fact, this will be the trickiest part of the proof).

Proof. Fix a, b, $c \geqslant 1$, and let $\mathcal{V}(a, b, c)$ denote the $\mathbf{F}_p$-subspace of $\mathbf{F}_p[A, B, C]$ of polynomials F such that

$$\deg_A(F) < a, \quad \deg_B(F) < b, \quad \deg_C(F) < c,$$

and similarly let $\mathcal{H}(d)$ be the subspace of polynomials of degree $< d$ in $\mathbf{F}_p[X]$. (These are analogue to Riemann-Roch spaces). Note that $\dim \mathcal{V}(a, b, c) = abc$ and $\dim \mathcal{H}(d) = d$.

We now observe that

$$\deg_A(\partial F) < a+1, \quad \deg_B(\partial F) < b+1, \quad \deg_C(\partial F) < c,$$

and hence, by immediate induction on $j \geqslant 0$, the map ∂^j is a linear map

$$\mathcal{V}(a, b, c) \xrightarrow{\partial^j} \mathcal{V}(a+j, b+j, c).$$

On the other hand, the map

$$\Psi \,:\, F \mapsto H = F(X, X, r)$$

sends $\mathcal{V}(a, b, c)$ to $\mathcal{H}(a+b)$, and hence we have linear maps

$$\Psi_j = \Psi \circ \partial^j \,:\, \mathcal{V}(a, b, c) \to \mathcal{H}(a+b+2j).$$

In these terms, our goal is to show that, under the stated conditions on a, b and c, there is a non-zero element in the kernel of the linear map T

$$\mathcal{V}(a, b, c) \xrightarrow{T} \bigoplus_{0 \leqslant j \leqslant m-1} \mathcal{H}(a+b+2j)$$

that sends F to

$$T(F) = (\Psi(F), \Psi_1(F), \ldots, \Psi_{m-1}(F)).$$

By linear algebra, this is true as soon as

$$\begin{aligned} abc = \dim \mathcal{V}(a,b,c) &> \dim \bigoplus_{0\leqslant j\leqslant m-1} \mathcal{H}(a+b+2j) \\ &= \sum_{0\leqslant j\leqslant m-1} (a+b+2j) \\ &= m(a+b+m-1), \end{aligned}$$

which is exactly what we stated. □

Now for the last step, which shows that the auxiliary polynomial G can also be taken non-zero, under suitable assumptions – this is the analogue of Lemma 4.28. Note that, up to this point, we have not used much information concerning the polynomial L_p.

Lemma 6.11 (Injectivity of specialization). *Let $m < p$ and let a, b, $c \geqslant 1$ be integers such that*

$$ab \leqslant p. \tag{6.9}$$

Then, with notation as in the proof of the previous lemma, the restriction of Φ given by

$$\Phi \begin{cases} \mathcal{V}(a,b,c) & \longrightarrow & \mathbf{F}_p[X] \\ F & \mapsto & F(X, X^p, L_p(X)) \end{cases}$$

is injective.

Before proving this last fact – which will be somewhat delicate –, we use it to finish the proof of Theorem 6.6.

Proof of Theorem 6.6. From our discussion (see, in particular, (6.6)), and the combination of Lemmas 6.10 and 6.11, we see that, whenever $m \geqslant 0$ and a, b, $c \geqslant 1$ are integers for which

$$\begin{cases} m < p \\ m(a+b+m-1) < abc \\ ab \leqslant p, \end{cases}$$

we have

$$|\mathcal{N}_r(\mathbf{F}_p)| \leqslant \frac{a+pb+(p-1)c}{m},$$

and we need to optimize this to obtain the smallest bound possible.

As in the end of Section 4.3, we explain how to reach the values of m, a, b and c that we will select. First, note that we need to have $m > b$, $m > c$ to beat the trivial bound on $|\mathcal{N}_r(\mathbf{F}_p)|$. Now, in terms of order of magnitude, if $m > a$, the second condition roughly becomes

$$m^2 < abc,$$

and since m should be as large as possible, this means that m should be about $\sqrt{abc}$; moreover, to have this as large as possible, we should have ab as large as possible, i.e., about $ab \approx p$. Since we assume $a < m$, the upper-bound on $|\mathcal{N}_r(\mathbf{F}_p)|$ is then

$$\approx \sqrt{p}\Big(\frac{b}{\sqrt{c}} + \sqrt{c}\Big),$$

and to minimize this one must take $b = c$, and b as small as possible, which means $m \approx \sqrt{pb}$ as small as possible, under the constraint $m > a$. This is done when there is equality, and so a, b and p are then related by relations

$$\begin{cases} ab \approx p \\ \sqrt{pb} \approx a, \end{cases}$$

which have solutions of size

$$a \approx p^{2/3}, \qquad b \approx p^{1/3}$$

from which we get the final guess

$$a \approx p^{2/3}, \qquad b \approx p^{1/3}, \qquad c = b \approx p^{1/3}, \qquad m = a \approx p^{2/3},$$

leading to an upper bound of type

$$|\mathcal{N}_r(\mathbf{F}_p)| \ll \sqrt{pb} = p^{2/3}.$$

It is easily checked, in a similar way, that assuming $m \leqslant a$ leads to the same solutions in terms of order of magnitude.

Now, to get exact constants as in (6.5), the following can be done: take

$$m = a = \lfloor p^{2/3} \rfloor, \qquad b = \lfloor p^{1/3} \rfloor,$$

and then check that, of course, $m < p$, $ab \leqslant p$, and that for

$$c = 10\lceil p^{1/3} \rceil,$$

we have[1]

$$m(a + b + m - 1) < abc$$

for all $p \geqslant 2$. The upper bound becomes

$$|\mathcal{N}_r(\mathbf{F}_p)| \leqslant 1 + p\frac{\lfloor p^{1/3} \rfloor}{\lfloor p^{2/3} \rfloor} + 10p\frac{\lceil p^{1/3} \rceil}{\lfloor p^{2/3} \rfloor},$$

and one checks (wastefully) that

$$p\frac{\lfloor p^{1/3} \rfloor}{\lfloor p^{2/3} \rfloor} \leqslant 3p^{2/3}, \quad p\frac{\lceil p^{1/3} \rceil}{\lfloor p^{2/3} \rfloor} \leqslant 4p^{2/3}$$

for all $p \geqslant 2$, which gives

$$|\mathcal{N}_r(\mathbf{F}_p)| \leqslant 43p^{2/3} + 1 \leqslant 44p^{2/3},$$

for $p \geqslant 2$ and $r \in \mathbf{F}_p$. □

And now for the final step: proving Lemma 6.11. The intuition is the following: if $\Phi(F) = 0$ with F of "small" degree, this means that the truncated logarithm L_p satisfies (essentially) a polynomial equation, i.e., that it behaves like an algebraic function. However, we know that, in $\mathbf{Q}[[X]]$, the element $\log(1/(1 - X))$ is *transcendental*. So we may expect that such relations are not possible.

Proof of Lemma 6.11. We first write F as a polynomial in B with coefficients in $\mathbf{F}_p[A, C]$, which gives a formula of the type

$$\Phi(F) = \sum_j F_j(X, L_p(X))X^{pj}.$$

[1] This is easier to do with a graphing software.

Now we observe that if $\Phi(F) = 0$, there exists some j such that $G_j = F_j(X, L_p(X))$ is non-zero but satisfies

$$v(G_j) \geqslant p$$

where $v(\cdot)$ denotes the order of vanishing of a polynomial at 0; indeed, if that is not the case, the different terms $F_j(X, L_p(X))X^{pj}$ satisfy

$$v(F_j(X, L_p(X))X^{pj}) = v(G_j) + pj,$$

and hence they are all distinct as j varies; but then the order of vanishing at 0 of $\Phi(F)$ is the minimal among these $v(G_j) + pj$, and in particular it is finite, which means that $\Phi(F) \neq 0$.

Now we are going to prove the following fact:

Fact 1. Assume $F \in \mathbf{F}_p[A, C]$ is not identically zero and satisfies

$$\deg_A(F) < a, \qquad \deg_C(F) < c,$$

and $ac \leqslant p$. Then $v(F(X, L_p(X))) < p$.

If we prove this, then obviously we are done. For the proof, we follow [22, §3], which is shorter than Heath-Brown's original argument. This proceeds by proving a more refined statement, which is better suited to a clever induction argument. Thus we will show:

Fact 2. Assume $F \in \mathbf{F}_p[A, C]$ is not zero and is of the form

$$F = \sum_{k<c} F_k C^k$$

where $F_k \in \mathbf{F}_p[A]$ has degree $\deg_A F_k \leqslant a_k$, $F_{c-1} \neq 0$, and

$$a_0 \geqslant a_1 \geqslant \cdots \geqslant a_{c-1}.$$

Denoting

$$d = a_0 + \cdots + a_{c-1}, \tag{6.10}$$

if $d + c - 1 < p$, we have $v(F(X, L_p(X))) \leqslant d + c - 1$.

In the situation of Fact 1, we can take $a_k = a - 1$ for $0 \leqslant k < c$, so that $d = (a-1)c$, and we have

$$d + c - 1 = ac - 1 < p$$

by assumption, so we deduce that $v(F(X, L_p(X))) \leqslant d + c - 1 = ac - 1 < p$, as claimed. So Fact 2 is indeed more general.

We will prove Fact 2 by induction on $c \geqslant 1$; observing that for $c = 1$, the result is obvious (F is then in $\mathbf{F}_p[A]$ so $F(X)$ vanishes to order $\leqslant \deg(F) \leqslant a_0 = d$), we assume that Fact 2 holds for c replaced by $c - 1$ or smaller integers, and we argue for a given $c \geqslant 2$ by induction on the quantity d given by (6.10). If $d = 0$, F is now in $\mathbf{F}_p[C]$, and the result follows from the fact that $v(L_p) = 1$ (i.e., 0 is a root of L_p with multiplicity 1) and F has degree $\leqslant c - 1$.

Thus assume that $d \geqslant 1$ and Fact 2 is valid when $a_0 + \cdots + a_{c-1} < d$ (and $a_0 + \cdots + a_{c-1} + c - 1 < p$). We suppose that $F \in \mathbf{F}_p[A, C]$ is given as above with $\deg_C(F) = c - 1$ and

$$d = a_0 + \cdots + a_{c-1},$$

and satisfies $v(\Phi(F)) = v(F(X, L_p(X))) \geqslant c + d$. To make use of the induction assumption, we want to consider the derivative of $G = \Phi(F)$; if we notice that the logarithm $\log(1/(1-X))$ satisfies the very simple relation

$$(X - 1)(\log(1/(1-X)))' = -1,$$

it is natural to consider $H = (X-1)G'$. Indeed, we have

$$(X-1)L_p' = (X-1)(1+X+\cdots+X^{p-2}) = X^{p-1}-1,$$

and hence for $k \geqslant 1$, we get

$$(X-1)(L_p^k)' = kL_p^{k-1}(X^{p-1}-1).$$

This means in particular that

$$\begin{aligned}(X-1)G' &\equiv \sum_{0\leqslant k<c}(X-1)F_k'L_p^k - \sum_{1\leqslant k<c} kF_kL_p^{k-1} \pmod{X^{p-1}}\\ &\equiv \Phi(H) \pmod{X^{p-1}}\end{aligned}$$

where

$$H = \sum_{0\leqslant k<c-1}\Bigl((A-1)F_k' - (k+1)F_{k+1}\Bigr)C^k \\ - (X-1)F_{c-1}C^{c-1} \in \mathbf{F}_p[A,C].$$

Because $v((X-1)G') \geqslant c+d-1$ and $c+d-1 \leqslant p-1$, the congruence implies that

$$v(\Phi(H)) \geqslant c+d-1.$$

Now we look at the coefficients of C^k in H, trying to apply the induction to it; since the a_i's are assumed to be non-increasing, we have

$$\deg_A((A-1)F_k' - (k+1)F_{k+1}) \leqslant a_k$$

for $0 \leqslant k < c-1$, and $\deg_A((A-1)F_{c-1}') \leqslant a_{c-1}$. Thus the quantity d has not decreased in H; however, if we consider instead

$$\tilde{H} = H - (\deg F_{c-1})F,$$

the coefficients of C^k for $k < c-1$ are still of degree $\leqslant a_k$, but the coefficient of C^{c-1} is now

$$(A-1)F_{c-1}' - (\deg F_{c-1})F_{c-1}$$

which has degree strictly smaller than F_{c-1}. Since we have

$$v(\Phi(\tilde{H})) \geqslant c+d-1,$$

we can apply the induction argument to $\tilde{H}$, and because $v(\Phi(\tilde{H})) \geqslant c+d-1$ contradicts the conclusion, the only possibility is that $\tilde{H} = 0$. But this means, in particular, that

$$\begin{cases}(A-1)F_{c-1}' - (\deg F_{c-1})F_{c-1} = 0\\ (A-1)F_{c-2}' - (\deg F_{c-1})F_{c-2} = -(c-1)F_{c-1}\end{cases}$$

The first relation implies that F_{c-1} is of the form

$$F_{c-1} = \alpha(A-1)^r$$

for some $r = \deg(F_{c-1}) \geqslant 0$ and $\alpha \neq 0$. Now we look at the terms of highest degree in the second relation; writing

$$F_{c-2} = \beta A^s + \cdots, \qquad s = \deg(F_{c-2}),$$

we find

$$(A-1)F_{c-2}' - (\deg F_{c-1})F_{c-2} = \beta(s-r)A^s + \cdots,$$

and this means that the second relation implies first $s = r$, and then the coefficient of $A^s = A^r$ must vanish, i.e., we must have

$$-(c-1)\alpha = 0,$$

which is not possible since $2 \leqslant c \leqslant p$. This shows that $F \neq 0$ can not exist, and finishes the induction. □

Bibliography

[1] E. Bombieri: *Counting points on curves over finite fields (d'après S. A. Stepanov)*, Séminaire N. Bourbaki, exposé no. 430, Lecture Notes in Math. 383 (1974), 234–241.

[2] J. Bourgain: *Mordell's exponential sum estimate revisited*, Journal A.M.S 18 (2005), 477–499.

[3] P. Deligne: *Cohomologie étale*, S.G.A $4\frac{1}{2}$, L.N.M 569, Springer Verlag (1977).

[4] P. Deligne: *La conjecture de Weil : I*, Publ. Math. IHÉS 43 (1974), 273–307

[5] P. Deligne: *La conjecture de Weil, II*, Publ. Math. IHÉS 52 (1980), 137–252.

[6] G.H. Hardy and E.M. Wright: *An introduction to the theory of numbers*, 5th Edition, Oxford Univ. Press, 1979.

[7] D.R. Heath-Brown and S.J. Patterson: *The distribution of Kummer sums at prime arguments*, J. reine angew. Math. 310 (1979), 111–130.

[8] D.R. Heath-Brown: *An estimate for Heilbronn's exponential sum*, in "Analytic number theory in honor of H. Halberstam", Birkhaüser (1996), 451–463.

[9] D.R. Heath-Brown and S. Konyagin: *New bounds for Gauss sums derived from k-th powers and for Heilbronn's exponential sum*, Quart. J. Math. 51 (2000), 221–235.

[10] K. Ireland and M. Rosen: *A Classical Introduction to Modern Number Theory*, 2nd Edition, GTM 84, Springer-Verlag (1990).

[11] H. Iwaniec: *Topics in classical automorphic forms*, Grad. Studies in Math. 17, A.M.S (1997).

[12] H. Iwaniec and E. Kowalski: *Analytic Number Theory*, A.M.S Colloq. Publ. 53, A.M.S (2004).

[13] N. Katz: *Gauss sums, Kloosterman sums and monodromy*, Annals of Math. Studies, 116, Princeton Univ. Press, 1988.

[14] E. Kowalski: *Exponential sums over finite fields, II: introduction to cohomological methods*, lectures notes from course at ETH Zürich, Fall Semester 2010, http://www.math.ethz.ch/~kowalski/exp-sumsII.pdf

[15] E. Kowalski: *Poincaré and analytic number theory*, in "The scientific legacy of Poincaré", É. Charpentier, É. Ghys, A. Lesne, editors, History of Math. 10, AMS 2010.

[16] E. Kowalski: these lecture notes and related documents, available online at http://www.math.ethz.ch/~kowalski/exp-sums.html

[17] D.A. Mit'kin: *Stepanov method of the estimation of the number of roots of some equations* (russian), Mater. Zametki 51 (1992), 52–58; English translation, Math. Notes 51 (1992), 565–570.

[18] W. Schmidt: *Equations over finite fields: an elementary approach*, Lecture Notes in Math. 536, Springer Verlag 1974.

[19] J. Silverman: *The arithmetic of elliptic curves*, Grad. Texts in Math. 106, Springer Verlag 1986.

[20] A. Weil: *Numbers of solutions of equations in finite fields*, Bull. A.M.S 55 (1949), 497–508.

[21] A. Weil: comments on [20], Collected Works, vol. I, 568–569, Springer 1979.

[22] H.B. Yu: *Note on Heath-Brown's estimate for Heilbronn's exponential sum*, Proc. American Math. Soc. 127 (1999), 1995–1998.