Analytic Number Theory

Henryk Iwaniec and Emmanuel Kowalski

Author addresses:

Rutgers University
E-mail address: iwaniec@math.rutgers.edu

ETH Zrich
E-mail address: kowalski@math.ethz.ch

1991 *Mathematics Subject Classification.* Primary 11Fxx, 11Lxx, 11Mxx, 11Nxx, 11T23, 11T24, 11R42.
Key words and phrases. Analytic number theory, distribution of
prime numbers, automorphic forms, L-functions, exponential sums.

CHAPTER 11

SUMS OVER FINITE FIELDS

11.1. Introduction.

In this chapter we consider a special type of exponential and character sums, called sometimes "complete sums", which can be seen as sums over the elements of a finite field. Although the methods of Chapter 8 can still be applied to the study of such sums, disregarding this special feature, the deepest understanding and the strongest results are obtained when the finite field aspect is taken into account and the powerful techniques of algebraic geometry are brought to bear.

We have already encountered in the previous chapters some examples of exponential sums which can be interpreted as sums over finite fields, for example, the quadratic Gauss sums

$$G_a(p) = \sum_{x \bmod p} \Big(\frac{x}{p}\Big) e\Big(\frac{ax}{p}\Big)$$

or the Kloosterman sums (1.56)

$$S(a,b;p) = \sideset{}{^\star}\sum_{x \bmod p} e\Big(\frac{ax+b\bar{x}}{p}\Big).$$

In this chapter we will study these sums in particular. The culminating point of our presentation is the elementary method of Stepanov which we apply for proving Weil's bound for Kloosterman sums

$$|S(a,b;p)| \leqslant 2\sqrt{p}$$

and Hasse's bound for the number of points of an elliptic curve over a finite field. Then we survey briefly, without proofs, the powerful formalism of ℓ-adic cohomology developed by Grothendieck, Deligne, Katz, Laumon and others, hoping to convey a flavor of the tools involved and to give the reader enough knowledge to make at least a preliminary analysis of any exponential sum he or she may encounter in analytic number theory.

11.2. Finite fields.

We first recall briefly some facts about finite fields, and establish the notations used in this chapter. For every prime p, the finite ring $\mathbb{Z}/p\mathbb{Z}$ of residue classes modulo p is a field, which we denote $\mathbb{F}_p$. The Galois theory of $\mathbb{F}_p$ is very easy to describe: for any $n \geqslant 1$, there exists a unique (up to isomorphism) field extension of $\mathbb{F}_p$ of degree n, written $\mathbb{F}_{p^n}$. Conversely, any finite field $\mathbb{F}$ with q elements is isomorphic (but not canonically) to a unique field $\mathbb{F}_{p^d}$, so $q = p^d$, and $\mathbb{F}$ admits also a unique finite extension of degree n for any $n \geqslant 1$, namely $\mathbb{F}_{p^{dn}}$.

Let now $\mathbb{F} = \mathbb{F}_q$ be a finite field with $q = p^d$ elements. In most of the chapter, p is fixed and we change notation slightly, denoting by $\overline{\mathbb{F}}$ an algebraic closure of $\mathbb{F}$

and by $\mathbb{F}_n \subset \overline{\mathbb{F}}$ the unique extension of degree n of $\mathbb{F}$ for $n \geqslant 1$. The context will always indicate clearly that the cardinality of $\mathbb{F}_n$ is q^n and not n.

The extension $\mathbb{F}_n/\mathbb{F}$ is a Galois extension, with Galois group G_n canonically isomorphic to $\mathbb{Z}/n\mathbb{Z}$, the isomorphism being the map $\mathbb{Z}/n\mathbb{Z} \to G_n$ defined by $1 \mapsto \sigma$, where σ is the Frobenius automorphism of $\mathbb{F}_n$ given by $\sigma(x) = x^q$.

Let $\bar{\mathbb{F}}$ be a given algebraic closure of $\mathbb{F}$, so by the above,

$$\bar{\mathbb{F}} = \bigcup_{n \geqslant 1} \mathbb{F}_n.$$

By Galois theory, for any $x \in \bar{\mathbb{F}}$, we have $x \in \mathbb{F} \iff \sigma(x) = x$ if and only if $x^q = x$ and more generally

$$x \in \mathbb{F}_n \iff \sigma^n(x) = x \iff x^{q^n} = x. \tag{11.1}$$

From this we can deduce that $\mathbb{F}_n$ is the splitting field of the polynomial $X^{q^n} - X \in \mathbb{F}[X]$. More precisely, one can state the following result of Gauss:

LEMMA 11.1. *For any integer $n \geqslant 1$, we have*

$$\prod_{d|n} \prod_{\deg(P)=d} P = X^{q^n} - X \tag{11.2}$$

where the product ranges over all irreducible monic polynomials P of degree d dividing n.

PROOF. This is an immediate consequence of the description of finite fields: the roots of the polynomial on the right side (in an algebraic closure) are exactly the elements $x \in \mathbb{F}_n$ with multiplicity one and, conversely, every such x has a minimal polynomial which must occur, exactly once, among the polynomials P on the left side. □

Associated to the extension $\mathbb{F}_n/\mathbb{F}$ are the trace map and the norm map. Because of the above description of the Galois group of $\mathbb{F}_n/\mathbb{F}$, the trace map $\mathrm{Tr} = \mathrm{Tr}_{\mathbb{F}_n/\mathbb{F}} : \mathbb{F}_n \to \mathbb{F}$ is given by

$$\mathrm{Tr}\,(x) = \sum_{0 \leqslant i \leqslant n-1} \sigma^i(x) = \sum_{0 \leqslant i \leqslant n-1} x^{q^i} \tag{11.3}$$

while the norm map $N = N_{\mathbb{F}_n/\mathbb{F}} : \mathbb{F}_n^* \to \mathbb{F}^*$ is similarly

$$N(x) = \prod_{0 \leqslant i \leqslant n-1} \sigma^i(x) = \prod_{0 \leqslant i \leqslant n-1} x^{q^i} = x^{\frac{q^n-1}{q-1}}. \tag{11.4}$$

The equations $\mathrm{Tr}\,(x) = y$ and $N(x) = y$, for a fixed $y \in \mathbb{F}$ are very important. Because the extension $\mathbb{F}_n/\mathbb{F}$ is separable, the equation $\mathrm{Tr}\,(x) = y$ always has a solution. If x_0 is a given solution, then all solutions are in one-to-one correspondence with solutions of $\mathrm{Tr}\,(a) = 0$, by $x = x_0 + a$. Moreover, any solution of $\mathrm{Tr}\,(a) = 0$ is of the form $a = \sigma(b) - b = b^q - b$ for some $b \in \mathbb{F}_n$, unique up to addition of an element in $\mathbb{F}$.

Similarly, for any $y \in \mathbb{F}^*$, the equation $N(x) = y$ has a solution, and if x_0 is a given solution, the set of solutions is in one-to-one correspondence with solutions of $N(a) = 1$, which by Hilbert's Theorem 90 (or by direct proof) are all given by

$a = \sigma(b)b^{-1} = b^{q-1}$ for some $b \in \mathbb{F}_n^*$, unique up to multiplication by an element in $\mathbb{F}^*$.

As the additive group of $\mathbb{F}$ is finite, the general theory of characters of a finite abelian group (see Chapter 3) can be applied. Characters of $\mathbb{F}$ are called additive characters, and they are all of the form $x \mapsto \psi(ax)$ for some $a \in \mathbb{F}$, where ψ is some fixed non-trivial additive character. For instance, let $\mathrm{Tr} : \mathbb{F} \to \mathbb{Z}/p\mathbb{Z}$ be the trace map to the base-field, then

$$\psi(x) = e(\mathrm{Tr}\,(x)/p) \tag{11.5}$$

is a non-trivial additive character of $\mathbb{F}$. For a given additive character ψ and $a \in \mathbb{F}$, we denote by ψ_a the character $x \mapsto \psi(ax)$.

Applying the general theory of characters of finite abelian groups, we get the orthogonality relations

$$\sum_{\psi} \psi(x) = \begin{cases} q & \text{if } x = 1, \\ 0 & \text{otherwise} \end{cases}$$

(which is used to "solve" the equation $x = 0$ in $\mathbb{F}$) and

$$\sum_{x \in \mathbb{F}} \psi(x) = \begin{cases} q & \text{if } \psi = 1 \text{ is the trivial character}, \\ 0 & \text{otherwise.} \end{cases}$$

The description of characters of the multiplicative group $\mathbb{F}^*$ (also called multiplicative characters of $\mathbb{F}$) is not so explicit. The group structure of $\mathbb{F}^*$ is well-known (dating back to Gauss): it is a cyclic group of order $q-1$. Generators of $\mathbb{F}^*$ are called primitive roots, and there are $\varphi(q-1)$ of them, but no useful formula for a primitive root exists. Fixing one, say $z \in \mathbb{F}^*$, one has an isomorphism

$$\log : \begin{cases} \mathbb{F}^* \simeq \mathbb{Z}/(q-1)\mathbb{Z} \\ x \mapsto n \text{ such that } z^n = x \end{cases}$$

and all multiplicative characters of $\mathbb{F}$ are expressed as

$$\chi(x) = e\Big(\frac{a \log(x)}{q-1}\Big)$$

for some $a \in \mathbb{Z}/(q-1)\mathbb{Z}$, but such a description is usually of no use in analytic number theory.

As examples of multiplicative characters, suppose $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$ and $p \neq 2$. Then the Legendre symbol

$$x \mapsto \Big(\frac{x}{p}\Big)$$

is a non-trivial quadratic character. In general, if $\delta \mid (q-1)$, there is a cyclic group of order δ consisting of characters χ of $\mathbb{F}^*$ of order δ.

The orthogonality relations become

$$\sum_{\chi} \chi(x) = \begin{cases} q-1 & \text{if } x = 1, \\ 0 & \text{otherwise,} \end{cases}$$

(the sum over all multiplicative characters), and

$$\sum_{x\in\mathbb{F}^*}\chi(x)=\begin{cases} q-1 & \text{if } \chi=1 \text{ is the trivial character,}\\ 0 & \text{otherwise.}\end{cases}$$

It is usual to extend multiplicative characters to $\mathbb{F}$ by defining $\chi(0)=0$ if $\chi\neq 1$, and $\chi(0)=1$ if $\chi=1$. Notice then that for any $\delta\mid q-1$ the formula

$$\sum_{\chi^\delta=1}\chi(x)=|\{y\in\mathbb{F}\mid y^\delta=x\}| \tag{11.6}$$

(also a particular case of the orthogonality relations for the group $\mathbb{F}^*/(\mathbb{F}^*)^d$, as described in Chapter 3) is true for all $x\in\mathbb{F}$.

11.3. Exponential sums.

Let $\mathbb{F}=\mathbb{F}_q$ be a finite field with $q=p^m$ elements, p a prime. Exponential sums over $\mathbb{F}$ can be of various kinds. For the simplest case, consider a polynomial $P\in\mathbb{F}[X]$ and an additive character ψ, and define the sum

$$S(P)=\sum_{x\in\mathbb{F}}\psi(P(x)).$$

Slightly more generally, take a non-zero rational function $f=P/Q\in\mathbb{F}(X)$ and consider

$$S(f)=\sum_{\substack{x\in\mathbb{F}\\ Q(x)\neq 0}}\psi(f(x));$$

for instance, taking $q=p$ and $f(x)=ax+bx^{-1}$, we have $S(f)=S(a,b;p)$. Multiplicative characters can also be used, getting sums of the type

$$S_\chi(f)=\sum_{x\in\mathbb{F}}^\star\chi(f(x))$$

(where the star in $\sum^\star$ means here and henceforth that the summation extends to all x which are not poles of f). For $q=p$, $\chi=\left(\frac{\cdot}{p}\right)$ (the Legendre symbol) and $f(x)\in\mathbb{Z}[X]$ a cubic polynomial without multiple roots modulo p, we see that $-S_\chi(f)$ is the p-th coefficient a_p of the Hasse-Weil zeta function of the elliptic curve with equation $y^2=f(x)$ (see Section 14.4).

Still more generally, one can mix additive and multiplicative characters, and define sums such as

$$S_\chi(f,g)=\sum_{x\in\mathbb{F}}^\star\chi(f(x))\psi(g(x)), \tag{11.7}$$

an example of which is the Salié sum $T(a,b;p)$ defined by

$$T(a,b;p)=\sum_{x \bmod p}^\star\Big(\frac{x}{p}\Big)e\Big(\frac{ax+b\bar{x}}{p}\Big)$$

which occurs in the Fourier expansion of half-integral weight modular forms; see [**I6**] for instance. In contrast with the seemingly simpler Kloosterman sums $S(a,b;p)$, the Salié sums $T(a,b;p)$ can be explicitly computed (see Lemma 12.4, and Corollary 21.9 for the uniform distribution of the "angles" of the Salié sums).

To end this list, we mention that all these definitions can again be generalized to sums in more than one variable, and that the summation variables can be restricted to the rational points of an algebraic variety defined over $\mathbb{F}$: some examples will appear in the survey sections of this chapter.

The exponential sums which directly arise in analytic number theory are sums over the prime field $\mathbb{Z}/p\mathbb{Z}$. However, the deeper understanding naturally requires considering sums over the extension fields $\mathbb{F}_{p^n}$. Indeed, the very reason for the success of algebraic methods lies in the fact that an exponential sum over $\mathbb{F}_p$ doesn't really come alone, but has natural "companions" over all the extension fields $\mathbb{F}_{p^n}$, and it is really the whole family which is investigated and which is the natural object of study. Those companion sums are easily defined: take the most general sum $S = S_\chi(f, g)$ we have introduced, then for $n \geqslant 1$ let

$$S_n = \sum_{x \in \mathbb{F}_n}^{\star} \chi(N_{\mathbb{F}_n/\mathbb{F}}(f(x)))\psi(\mathrm{Tr}_{\mathbb{F}_n/\mathbb{F}}(g(x))) \tag{11.8}$$

where we use the multiplicative character $\chi \circ N$ and the additive character $\psi \circ \mathrm{Tr}$ of $\mathbb{F}_n$. All the sums S_n are incorporated into a single object, the zeta function of the exponential sum, which is the formal power series $Z = Z_\chi(f, g) \in \mathbb{C}[[T]]$ defined by the formula

$$Z = \exp\Big(\sum_{n \geqslant 1} \frac{S_n}{n} T^n\Big).$$

Justification for the introduction of the zeta function comes from the following rationality theorem, conjectured by Weil, and proved by Dwork.

THEOREM 11.2 (DWORK). *The zeta function Z is the power series expansion of a rational function; more precisely, there exist coprime polynomials P and Q in $\mathbb{C}[T]$, with $P(0) = Q(0) = 1$, such that $Z = \frac{P}{Q}$.*

As a corollary, denote by (α_i) (resp. (β_j)) the inverse of the roots (with multiplicity) of P (resp. Q), so

$$P = \prod_i (1 - \alpha_i T), \quad Q = \prod_j (1 - \beta_j T).$$

Then using the power-series expansion

$$\log \frac{1}{1 - T} = \sum_{n \geqslant 1} \frac{T^n}{n}$$

we find that the formula $Z = P/Q$ is equivalent to the formula

$$S_n = \sum_j \beta_j^n - \sum_i \alpha_i^n$$

for any $n \geqslant 1$, which shows how the various sums S_n are related. In particular, note that they satisfy a linear recurrence relation of order d equal to the number $\deg P + \deg Q$ of roots α_i, β_j.

COROLLARY 11.3. *We have for any $n \geqslant 1$ the upper bound*

$$|S_n| \leqslant \sum_j |\beta_j|^n + \sum_i |\alpha_i|^n.$$

In particular,

$$|S| \leqslant \sum_j |\beta_j| + \sum_i |\alpha_i|. \tag{11.9}$$

A common abuse of language is to speak of the α_i and β_j as the roots of the exponential sum S. We will describe in Section 11.11 a number of general facts about these roots. In the intervening sections, we will prove Dwork's Theorem and estimate the modulus of the roots in the important special cases of Gauss sums, Kloosterman sums and for the local zeta function of elliptic curves.

REMARKS. We have called the sums S_n, $n \geqslant 1$, "companions" of the original exponential sum S. However, one can consider other companions of S as well. If S involves an additive character, it can also be very useful sometimes to consider S as just one element of the family of sums obtained by varying the additive character, specifically if

$$S = \sum_{x\in\mathbb{F}}^{\star} \chi(f(x))\psi(g(x)),$$

we also introduce for $a \in \mathbb{F}$,

$$S_a = \sum_{x\in\mathbb{F}}^{\star} \chi(f(x))\psi_a(g(x)) = \sum_{x\in\mathbb{F}}^{\star} \chi(f(x))\psi(ag(x)).$$

Estimates on average over a for the first few power moments of S_a are often easily derived by elementary means, and they can be of great use in estimating S, even in addition to the methods of algebraic geometry. See the proof of Weil's bound for Kloosterman sums in Section 11.7 and the examples in Section 11.11. More general types of families have been (and still are) extensively studied by Katz; see for instance [**K1**].

11.4. The Hasse-Davenport relation.

We consider general Gauss sums over a finite field. Let $\mathbb{F} = \mathbb{F}_q$ be a finite field with $q = p^m$ elements, and let ψ be an additive character and χ a multiplicative character of $\mathbb{F}$. The Gauss sum $G(\chi, \psi)$ is

$$G(\chi, \psi) = \sum_{x\in\mathbb{F}} \chi(x)\psi(x). \tag{11.10}$$

(recall that χ is extended to $\mathbb{F}$ by $\chi(0) = 1, 0$ according to whether χ is trivial or not). When χ is the Legendre symbol, one recovers quadratic Gauss sums.

The associated sums over the extensions fields are

$$G_n(\chi, \psi) = \sum_{x\in\mathbb{F}_n} \chi(N_{\mathbb{F}_n/\mathbb{F}}(x))\psi(\mathrm{Tr}_{\mathbb{F}_n/\mathbb{F}}(x))$$

and the zeta function is

$$Z(\chi, \psi) = \exp\Big(\sum_{n\geqslant 1} \frac{G_n(\chi, \psi)}{n} T^n\Big). \tag{11.11}$$

In this case, Dwork's Theorem was proved by Hasse and Davenport and is known as the Hasse-Davenport Relation.

THEOREM 11.4 (HASSE-DAVENPORT). *Assume χ and ψ are non-trivial. Then we have for any $n \geqslant 1$,*

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

or equivalently the zeta function is a linear polynomial

$$Z(\chi,\psi) = 1 + G(\chi,\psi)T.$$

Hence the only "root" for the Gauss sum is $G(\chi,\psi)$ itself. This can be estimated elementarily, as was done for the Gauss sums considered in Chapter 3.

PROPOSITION 11.5. *We have*

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

if neither χ nor ψ is trivial, while

$$|G(1,\psi)| = \begin{cases} 0 & \text{if } \psi \text{ non-trivial,} \\ q & \text{if } \psi = 1 \end{cases}$$

$$|G(\chi,1)| = \begin{cases} 0 & \text{if } \chi \text{ non-trivial,} \\ q & \text{if } \chi = 1. \end{cases}$$

PROOF. The last two statements are immediate, so assume neither χ nor ψ is trivial. We have

$$\begin{aligned} |G(\chi,\psi)|^2 &= \sum_{x,y\in\mathbb{F}^*} \chi(x)\bar{\chi}(y)\psi(x)\psi(-y) \\ &= \sum_{z\in\mathbb{F}} \chi(z) \sum_{y\in\mathbb{F}^*} \psi((z-1)y) \text{ (on writing } z = xy^{-1}) \\ &= q \text{ (by orthogonality, applied twice.)} \end{aligned}$$

□

We now turn to the proof of the Hasse-Davenport Relation. We consider the field $F = \mathbb{F}(X)$ of rational functions on $\mathbb{F}$ and the ring $R = \mathbb{F}[X]$ of polynomials. Recall that R is a principal ideal domain. For $h \in R$ of degree $d \geqslant 0$, we define the norm

$$N(h) = q^d.$$

The zeta function of F is the Dirichlet series (analogous to the Riemann zeta function)

$$\zeta_F(s) = \sum_{\substack{h\in R \\ h \text{ monic}}} N(h)^{-s}.$$

REMARK. This could also be written as a sum over the non-zero ideals $\mathfrak{a}$ in R,

$$\zeta_F(s) = \sum_{\mathfrak{a}} N(\mathfrak{a})^{-s}$$

where $N(\mathfrak{a}) = |R/\mathfrak{a}| = N(h)$ for any polynomial h such that $\mathfrak{a} = (h)$. But we will work with polynomials to emphasize the elementary spirit here.

The Dirichlet series $\zeta_F(s)$ converges absolutely for $\mathrm{Re}\,(s) > 1$. Indeed, putting

$$n(d) = \{h \in R \mid \deg(h) = d,\ h \text{ is monic}\} = q^d$$

we obtain immediately

$$\zeta_F(s) = \sum_{d \geqslant 0} n(d) q^{-ds} = \sum_{d \geqslant 0} q^{(1-s)d} = (1 - q^{1-s})^{-1}.$$

On the other hand, unique factorization into irreducible polynomials yields an expression of ζ_F as an Euler product

$$\zeta_F(s) = \prod_{\substack{P \in R \\ P \text{ monic irreducible}}} (1 - N(P)^{-s})^{-1}$$

which is convergent for $\mathrm{Re}\,(s) > 1$.

The first step in the proof of the Hasse-Davenport Relation consists of writing the zeta function of Gauss sums as an L-function for the field F. Let $H \subset F^*$ be the subgroup of rational functions which are quotients of monic polynomials, and $G \subset H$ a subgroup with the property

$$h_1 h_2 \in G \Rightarrow h_1,\, h_2 \in G.$$

Then if $\alpha : G \to \mathbb{C}^*$ is a character of the group G, it can be extended to a totally multiplicative function of the set of monic polynomials $h \in R$ by putting $\alpha(h) = 0$ if $h \notin G$. The corresponding L-function is defined analogously to the classical L-functions by the Dirichlet series

$$L(s, \alpha) = \sum_{\substack{h \in R \\ h \text{ monic}}} \alpha(h) N(h)^{-s} = \prod_P (1 - \alpha(P) N(P)^{-s})^{-1}$$

for $\mathrm{Re}\,(s) > 1$.

For dealing with Gauss sums we consider the subgroup $G \subset H$ of rational functions f defined and non-vanishing at 0. Define a character λ on G by

$$\lambda(h) = \chi(a_d)\psi(a_1)$$

for $h = X^d - a_1 X^{d-1} + \cdots + (-1)^d a_d \in R$. Clearly λ is multiplicative on monic polynomials, and extends to a character of G. In this case we get the following:

LEMMA 11.6. *We have* $L(s, \lambda) = 1 + G(\chi, \psi) q^{-s}$.

PROOF. We arrange the Dirichlet series for $L(s, \lambda)$ according to the degree of h:

$$L(s, \lambda) = \sum_{d \geqslant 0} \Big(\sum_{\deg(h) = d} \lambda(h) \Big) q^{-ds}$$

and evaluate each term in turn. For $d = 0$, the only monic polynomial occurring is $h = 1$, and $\lambda(1) = 1$. For $d = 1$, we have $h = X - a$ so that

$$\sum_{\deg(h)=1} \lambda(h) = \sum_{a \in \mathbb{F}} \lambda(X - a) = \sum_{a \in \mathbb{F}} \chi(a)\psi(a) = G(\chi, \psi).$$

For any $d \geqslant 2$ we have

$$\sum_{\deg(h)=d} \lambda(h) = \sum_{a_1,\dots,a_d \in \mathbb{F}} \lambda(X^d - a_1 X^{d-1} + \cdots + (-1)^d a_d)$$
$$= q^{d-2} \sum_{a_1, a_d \in \mathbb{F}} \chi(a_d)\psi(a_1) = 0$$

by orthogonality, because at least one of the characters χ, ψ is non-trivial. □

On the other hand, appealing to the Euler product we will prove:

LEMMA 11.7. *We have $L(s,\lambda) = Z(q^{-s})$, where $Z = Z(\chi,\psi)$ is the zeta function (11.11) associated with Gauss sums.*

Theorem 11.4 follows from Lemmas 11.6 and 11.7.

PROOF OF LEMMA 11.7. Taking the logarithmic derivative of the Euler product, we get

$$-\frac{1}{\log q}\frac{L'(s,\lambda)}{L(s,\lambda)} = \sum_P \deg(P) \sum_{r\geqslant 1} \lambda(P)^r q^{-rds}$$
$$= \sum_{n\geqslant 1} \Big(\sum_{rd=n} d \sum_{\substack{P \\ \deg(P)=d}} d\lambda(P)^r \Big) q^{-ns}$$

while, on the other hand,

$$-\frac{1}{\log q}\frac{Z'(q^{-s})}{Z(q^{-s})} = \sum_{n\geqslant 1} G_n(\chi,\psi) q^{-ns}.$$

It therefore suffices to prove the formula

$$\sum_{\substack{P \\ d=\deg(P)\mid n}} d\lambda(P)^{n/d} = G_n(\chi,\psi) \tag{11.12}$$

for $n \geqslant 1$, the equality of the logarithmic derivatives being sufficient to imply Lemma 11.7 since both sides are Dirichlet series with leading coefficient 1.

To prove (11.12), let P be one of the irreducible polynomials appearing on the left side, of degree $d \mid n$. Its roots, say x_1,..., x_d, are in $\mathbb{F}_n$. Fix one root $x = x_j$ and write

$$P = X^d - a_1 X^{d-1} + \cdots + (-1)^d a_d.$$

We get

$$N(x) = (N_{\mathbb{F}_d/\mathbb{F}}(x))^{n/d} = a_d^{n/d},$$
$$\mathrm{Tr}\,(x) = \frac{n}{d}\mathrm{Tr}_{\mathbb{F}_d/\mathbb{F}}(x) = \frac{n}{d}a_1$$

hence

$$\lambda(P)^{n/d} = (\chi(a_d)\psi(a_1))^{n/d} = \chi(a_d^{n/d})\psi\Big(\frac{n}{d}a_1\Big) = \chi(N(x))\psi(\mathrm{Tr}\,(x)).$$

Summing over all roots of P we derive

$$d\lambda(P)^{n/d} = \sum_{i=1}^{d} \chi(N(x_i))\psi(\mathrm{Tr}\,(x_i)),$$

and summing over all P with $\deg(P) \mid n$, we get (11.12) by Lemma 11.1 since every element in $\mathbb{F}_n$ will appear exactly once as one of the roots x_i for some P. □

11.5. The zeta function for Kloosterman sums.

Next we consider Kloosterman sums. Let $\mathbb{F}$ be a finite field with $q = p^m$ elements and this time consider additive characters ψ and φ. We define the Kloosterman sum associated to ψ and φ by

$$S(\psi, \varphi) = -\sum_{x \in \mathbb{F}^*} \psi(x)\varphi(x^{-1}), \tag{11.13}$$

(the minus factor is only for cosmetic reasons). When $q = p$ is prime, and $\psi(x) = e(ax/p)$, $\varphi(x) = e(bx/p)$, we have therefore $S(\psi, \varphi) = -S(a, b; p)$.

The companion sums over the extension fields $\mathbb{F}_n$ are

$$S_n(\psi, \varphi) = -\sum_{x \in \mathbb{F}_n^*} \psi(\mathrm{Tr}\,(x))\varphi(\mathrm{Tr}\,(x^{-1}))$$

and the Kloosterman zeta function is

$$Z = Z(\psi, \varphi) = \exp\Big(\sum_{n \geqslant 1} \frac{S_n(\psi, \varphi)}{n} T^n\Big).$$

We will prove Dwork's Theorem in this case, which is due to Carlitz.

THEOREM 11.8. *Assume that ψ and φ are both non-trivial. Then*

$$Z(\psi, \varphi) = \frac{1}{1 - S(\psi, \varphi)T + qT^2}.$$

The proof is very similar to that of Theorem 11.4. We put $R = \mathbb{F}[X]$, $F = \mathbb{F}(X)$ as before, and consider the same group $G \subset F^*$ of quotients of monic polynomials defined and non-vanishing at 0. We define a character $\eta : G \to \mathbb{C}^*$ by putting

$$\eta(h) = \psi(a_1)\varphi(a_{d-1}/a_d)$$

for a monic polynomial $h \in G$, where we write (compare the previous section)

$$h = X^d + a_1 X^{d-1} + \cdots + a_{d-1}X + a_d$$

(with $a_d \neq 0$ since $h \in G$). The following computation verifies that η is indeed a character of G: let $h' = X^e + b_1 X^{e-1} + \cdots + b_{e-1}X + b_e$ with $b_e \neq 0$, then

$$hh' = X^{d+e} + (a_1 + b_1)X^{d+e-1} + \cdots + (a_{d-1}b_e + a_d b_{e-1})X + a_d b_e$$

and

$$\begin{aligned} \eta(hh') &= \psi(a_1 + b_1)\varphi\Big(\frac{a_{d-1}b_e + a_d b_{e-1}}{a_d b_e}\Big) \\ &= \psi(a_1)\varphi(a_{d-1}/a_d)\psi(b_1)\varphi(b_{e-1}/b_e) \\ &= \eta(h)\eta(h'). \end{aligned}$$

Recall that we extend η to all $h \in R$ by putting $\eta(h) = 0$ for $h \notin G$.

LEMMA 11.9. *For ψ and ψ non-trivial, the L-function associated to η is given by*

$$L(s,\eta) = \sum_h \eta(h)N(h)^{-s} = 1 - S(\psi,\varphi)q^{-s} + q^{1-2s}.$$

PROOF. By arranging terms according to the degree of h, we write

$$L(s,\eta) = \sum_{d\geqslant 0}\Big(\sum_{\deg(h)=d}\eta(h)\Big)q^{-ds}$$

and evaluate the inner sums. For $d = 0$, we have only $h = 1$ and $\eta(1) = 1$. For $d = 1$, we have $h = X + a$ with $a \neq 0$, hence

$$\sum_{\deg(h)=1}\eta(h) = \sum_{a\in\mathbb{F}^*}\eta(X+a) = \sum_{a\in\mathbb{F}^*}\psi(a)\varphi(a^{-1}) = -S(\psi,\varphi).$$

For $d = 2$, we get

$$\begin{aligned}\sum_{\deg(h)=2}\eta(h) &= \sum_{\substack{a\in\mathbb{F}\\ b\in\mathbb{F}^*}}\eta(X^2+aX+b) = \sum_{\substack{a\in\mathbb{F}\\ b\in\mathbb{F}^*}}\psi(a)\varphi(ab^{-1})\\ &= q - 1 + \Big(\sum_{a\in\mathbb{F}^*}\psi(a)\Big)\Big(\sum_{b\in\mathbb{F}^*}\varphi(b)\Big) = q\end{aligned}$$

by applying twice the orthogonality of characters, since neither ψ nor φ is trivial.

Finally, for $d \geqslant 3$, we get

$$\begin{aligned}\sum_{\deg(h)=d}\eta(h) &= \sum_{a\in\mathbb{F}^*}\sum_{a_1,\dots,a_{d-1}\in\mathbb{F}}\eta(X^d + a_1X^{d-1} + \cdots + a_{d-1}X + a)\\ &= q^{d-3}\sum_{\substack{a_1,a_{d-1}\in\mathbb{F}\\ a\in\mathbb{F}^*}}\psi(a_1)\varphi(a_{d-1}a^{-1}) = 0\end{aligned}$$

since there is free summation over $a_1 \in \mathbb{F}$. □

LEMMA 11.10. *For ψ and φ non-trivial, we have the identity*

$$Z(\psi,\varphi)(q^{-s}) = L(s,\eta)^{-1} = \frac{1}{1 - S(\psi,\varphi)q^{-s} + q^{1-2s}}.$$

This lemma completes the proof of Theorem 11.8.

PROOF. The L-function has an Euler product

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

Taking the logarithmic derivative we get

$$\begin{aligned}-\frac{1}{\log q}\frac{L'(s,\eta)}{L(s,\eta)} &= \sum_P \deg(P)\sum_{r\geqslant 1}\eta(P)^r q^{-r\deg(P)s}\\ &= \sum_{n\geqslant 1}\Big(\sum_{rd=n} d\sum_{\deg(P)=r}\eta(P)^r\Big)q^{-ns}\end{aligned}$$

and as before it suffices to prove the formula

$$\sum_{d=\deg(P)|n} d\eta(P)^{n/d} = -S_n(\psi,\varphi) \tag{11.14}$$

for $n \geqslant 1$. Let

$$P = X^d + a_1X^{d-1} + \cdots + a_{d-1}X + a_d$$

be one of the irreducible polynomials on the left side of (11.14), of degree $d \mid n$, and x_1,..., x_d its roots, which lie in $\mathbb{F}_d$. We have for each i,

$$\mathrm{Tr}\,(x_i) = \frac{n}{d}\mathrm{Tr}_{\mathbb{F}_d/\mathbb{F}}(x_i) = -\frac{n}{d}a_1$$

(since $a_d^{-1}X^dP(X^{-1}) = X^d + \frac{a_{d-1}}{a_d}X^{d-1} + \cdots + a_d^{-1}$) and

$$\mathrm{Tr}\,(x_i^{-1}) = \frac{n}{d}\mathrm{Tr}_{\mathbb{F}_d/\mathbb{F}}(x_i^{-1}) = -\frac{n}{d}\frac{a_{d-1}}{a_d}.$$

Hence

$$\eta(P)^{n/d} = \psi\Big(\frac{n}{d}a_1\Big)\varphi\Big(\frac{n}{d}\frac{a_{d-1}}{a_d}\Big) = \psi(-x_i)\varphi(-x_i^{-1})$$

and summing over the roots x_i, then over the polynomials P of degree $d \mid n$, we obtain (11.14) by Gauss's Lemma again. □

Theorem 11.8 allows us to factor the Kloosterman zeta function

$$Z(\psi,\varphi) = (1 - S(\psi,\varphi)T + qT^2)^{-1} = (1-\alpha T)^{-1}(1-\beta T)^{-1}$$

where α and β are complex numbers, and of course $\alpha+\beta = S(\psi,\varphi)$, $\alpha\beta = q$. In sharp contrast to the case of Gauss sums, however, the roots α and β cannot be explicitly computed.

THEOREM 11.11 (WEIL). *Assume that ψ and φ are non-trivial and $p \neq 2$. Then the roots α and β for the Kloosterman sum $S(\psi,\varphi)$ satisfy $|\alpha| = |\beta| = \sqrt{q}$, and therefore we have*

$$|S(\psi,\varphi)| \leqslant 2\sqrt{q}. \tag{11.15}$$

We will prove Theorem 11.11 in the next two sections.

COROLLARY 11.12. *Let a, b, c be integers, c positive. We have*

$$|S(a,b;c)| \leqslant \tau(c)(a,b,c)^{1/2}c^{1/2}. \tag{11.16}$$

PROOF. By the twisted multiplicativity (1.59) for Kloosterman sums, it suffices to consider $c = p^\nu$ with p prime and $\nu \geqslant 1$. If $p \mid ab$, we have Ramanujan sums for which the result is easy (see (3.2), (3.3)). Otherwise, the case $\nu = 1$ follows from Theorem 11.11 for $p \geqslant 3$, and for $p = 2$ one checks immediately that the Kloosterman sums modulo 2 satisfy Theorem 11.11: we have $S(1,1;2) = 1$, and the associated zeta function is therefore $Z(T) = 1+T+2T^2$, with roots $(-1\pm i\sqrt{7})/4$ of modulus $1/\sqrt{2}$.

The case $p \nmid ab$ and $\beta \geqslant 2$ can be dealt with elementarily; see Exercise 1 of Chapter 12. □

EXERCISE 1. Consider a general Kloosterman-Salié sum

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

and its associated companions S_n and zeta function Z, where ψ and φ are additive characters of $\mathbb{F}$ and χ is multiplicative (so $\chi = 1$ is the case of Kloosterman sums). Show that

$$Z = (1 - S(\chi;\psi,\varphi)q^{-s} + \bar{\chi}(-a)\chi(b)q^{1-2s})^{-1}$$

where

$$\psi(x) = e\Big(\frac{\mathrm{Tr}\,(ax)}{p}\Big), \quad \varphi(x) = e\Big(\frac{\mathrm{Tr}\,(bx)}{p}\Big).$$

11.6. Stepanov's method for hyperelliptic curves.

We will prove Theorem 11.11 by deducing it from the Riemann Hypothesis for certain algebraic curves over finite fields. However, we use Stepanov's elementary method (see [**Ste**], [**Sch**], [**Bo3**]) instead of Weil's arguments.

Let $\mathbb{F}$ be a finite field with q element, of characteristic p. We will only consider algebraic curves C_f over $\mathbb{F}$ given by equations of the type

$$C_f \,:\, y^2 = f(x) \tag{11.17}$$

for some polynomial $f \in \mathbb{F}[X]$ of degree $m \geqslant 3$. We assume moreover the following condition

(11.18) The polynomial $Y^2 - f(X) \in \mathbb{F}[X,Y]$ is absolutely irreducible

(i.e. it is irreducible over the algebraic closure of $\mathbb{F}$). This is a minimal regularity assumption on the curve C_f. It is easily seen to be equivalent to the condition that f is not a square in $\bar{\mathbb{F}}[X]$, and we will use it in this form.

REMARK. Stepanov's method has been refined by Schmidt [**Sch**] and Bombieri [**Bo3**] and is capable of handling the general case of the Riemann Hypothesis for curves; the case of curves with equation of the type $y^d = f(x)$ is not much harder than the one treated here. We limit ourselves to the curves C_f for simplicity, and because it suffices for the application to Kloosterman sums and elliptic curves. Note that curves of the type $y^2 = f(x)$ are instances of so-called hyperelliptic curves, which are quite naturally distinguished among algebraic curves (but not all hyperelliptic curves are of this form; see for instance elliptic curves in characteristics 2 and 3).

The problem we consider is that of estimating the number $|C_f(\mathbb{F})|$ of $\mathbb{F}$-rational points of C_f, i.e., the number N of solutions $(x,y) \in \mathbb{F}^2$ to the equation $y^2 = f(x)$. We are especially interested in this question when q is large (typically, as with exponential sums, the polynomial $f \in \mathbb{F}[X]$ is fixed, and we consider the $\mathbb{F}_n$-rational points for all $n \geqslant 1$), although we will obtain completely explicit inequalities.

THEOREM 11.13. *Assume that $f \in \mathbb{F}[X]$ satisfies (11.18), and $m = \deg(f) \geqslant 3$. If $q > 4m^2$, then $N = |C_f(\mathbb{F})|$ satisfies*

$$|N - q| < 8m\sqrt{q}.$$

Clearly we can assume that $p > 2$, as otherwise the map $y \mapsto y^2$ is an automorphism of $\mathbb{F}$ and $N = q$.

Stepanov's idea, which was inspired by results of Thue [**Thu**] in diophantine approximation, is to construct an auxiliary polynomial of degree r, say, having zeros of high multiplicity (at least ℓ, say) at the x-coordinates of points of $C_f(\mathbb{F})$. Hence one gets easily the inequality

$$N \leqslant 2r\ell^{-1},$$

the factor two being the highest possible multiplicity of a given x-coordinate among points in $C_f(\mathbb{F})$. This inequality turns out to be so strong that it gives the upper-bound of the theorem (certainly a surprising fact!). A trick then deduces the lower-bound from this.

We first distinguish among the points (x, y) these with $y = 0$. Let N_0 be the number of distinct zeros of f in $\mathbb{F}$, which is also the number of points $(x, 0) \in C_f(\mathbb{F})$. If (x, y) is a point of C_f with $y \neq 0$, it follows that $f(x)$ is a square in $\mathbb{F}$, which is true if and only if $g(x) = 1$ where

$$g = f^c \quad \text{with} \quad c = \tfrac{1}{2}(q-1).$$

Conversely, given $x \in \mathbb{F}$ with $g(x) = 1$, there are exactly two elements $y \in \mathbb{F}^*$ with $y^2 = f(x)$. Hence, writing

$$N_1 = |\{x \in \mathbb{F} \mid g(x) = 1\}| \tag{11.19}$$

it follows that

$$N = N_0 + 2N_1. \tag{11.20}$$

We will estimate N_1 by following the strategy sketched above, but in order to handle the lower bound later, we generalize slightly and consider for any $a \in \mathbb{F}$ the set

$$\mathcal{S}_a = \{x \in \mathbb{F} \mid f(x) = 0 \text{ or } g(x) = a\}. \tag{11.21}$$

To produce polynomials vanishing to a large order, we wish to use derivatives to characterize when this occurs. In characteristic 0, a polynomial P has a zero of order ℓ at x_0 if and only if all the derivatives $P^{(i)}$ with $0 \leqslant i < \ell$ vanish at x_0. In characteristic $p > 0$, however, this is no longer true if $\ell > p$, as the example of the polynomial $P = X^p$ shows, since $P^{(k)} = 0$ for all $k \geqslant 1$, in particular, $P^{(p)}(0) = 0$. A satisfactory solution follows by considering other differential operators.

Definition. Let K be any field. For any $k \geqslant 0$, the k-th Hasse derivative is the linear operator $E^k : K[X] \to K[X]$ defined by

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

for all $n \geqslant 0$, and extended to $K[X]$ by linearity. We also write $E = E^1$ (but beware that $E^k \neq E \circ E \circ \cdots \circ E$).

Remark. From the binomial expansion

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

and by linearity, we see that the value of $E^k P$ at a point $a \in K$, for $P \in K[X]$, is simply the coefficient of $(X-a)^k$ in the Taylor expansion of P around a. This

explains the properties of the Hasse derivatives, but we cannot take this as a definition, because the values of a polynomial over a finite field do not characterize the polynomial.

Note that for K of characteristic $p > 0$, we get $EX^p = E^2X^p = \cdots = E^{p-1}X^p = 0$, but $E^pX^p = 1 \neq 0$ and we see that the Hasse derivatives detect the zero of X^p of order exactly p at 0. This is a general fact, as Lemma 11.16 will show.

LEMMA 11.14. *The Hasse derivatives satisfy*

$$E^k(fg) = \sum_{j=0}^{k} (E^j f)(E^{k-j} g)$$

for all f, $g \in K[X]$, and more generally,

$$E^k(f_1 \cdots f_r) = \sum_{j_1+\cdots+j_r=k} (E^{j_1} f_1) \cdots (E^{j_r} f_r) \tag{11.22}$$

for f_1, ..., $f_r \in K[X]$.

PROOF. It suffices to consider $f = X^m$, $g = X^n$, and the first formula follows from the identity

$$\binom{n+m}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$$

which is obvious from the combinatorial interpretation of the binomial coefficients. Then the second formula follows by induction. □

COROLLARY 11.15. *(1) For all k, $r \geqslant 0$, and all $a \in K$, we have*

$$E^k(X-a)^r = \binom{r}{k}(X-a)^{r-k}.$$

(2) For all k, $r \geqslant 0$ with $k \leqslant r$, and all f, $g \in K[X]$, we have

$$E^k(fg^r) = hg^{r-k}$$

for some polynomial h such that

$$\deg(h) \leqslant \deg(f) + k\deg(g) - k.$$

PROOF. For (1), we apply (11.22) to $f_1 = \cdots = f_r = X - a$, getting

$$E^k(X-a)^r = \sum_{j_1+\cdots+j_r=k} E^{j_1}(X-a) \cdots E^{j_r}(X-a)$$

and only terms with all $j_i \in \{0, 1\}$, $1 \leqslant i \leqslant r$, give non-zero contributions since $E^j(X-a) = 0$ for $j \geqslant 2$, from the definition. Hence (1) follows.

For (2), we observe that if $k \leqslant r$, we have $j_i = 0$ for at least $r - k$ indices in (11.22), which gives (2). □

LEMMA 11.16. *Let $f \in K[X]$ and $a \in K$. Suppose that $(E^k f)(a) = 0$ for all $k < \ell$. Then f has a zero of order $\geqslant \ell$ at a, i.e., is divisible by $(X-a)^\ell$.*

PROOF. Let

$$f = \sum_{0 \leqslant i \leqslant d} \alpha_i (X-a)^i$$

be the Taylor expansion of f around a. By (1) of Corollary 11.15, we obtain

$$E^k f = \sum_{k \leqslant i \leqslant d} \alpha_i \binom{i}{k} (X-a)^{i-k}$$

and evaluating at a we get $\alpha_k = 0$ for all $k < \ell$, hence f is divisible by $(X-a)^\ell$ as claimed. □

We need another technical lemma.

LEMMA 11.17. *Let $K = \mathbb{F}$ be a finite field of characteristic p with q elements, and let $r = h(X, X^q) \in \mathbb{F}[X]$, where $h \in \mathbb{F}[X, Y]$. Then*

$$E^k r = (E_X^k h)(X, X^q)$$

for all $k < q$, where on the right side $E_X^k h$ denotes the Hasse derivative of h performed with respect to X.

PROOF. It suffices to consider $h = X^n Y^m$, so we must prove that $E^k X^{n+mq} = (E^k X^n) X^{mq}$. From Lemma 11.14, we get

$$E^k X^{n+mq} = \sum_{j=0}^{k} E^{k-j} X^n E^j X^{mq}$$

so it suffices to show that $E^j X^{mq} = 0$ for $0 < j < q$ to prove the lemma. But

$$\binom{mq}{j} = \frac{mq}{j} \binom{mq-1}{j-1} = 0$$

in characteristic p, and the result follows. □

We come to the heart of Stepanov's method, the construction of the auxiliary polynomial.

PROPOSITION 11.18. *Assume that $q > 8m$, and let ℓ be an integer satisfying $m < \ell \leqslant q/8$. Then there exists a polynomial $r \in \mathbb{F}[X]$ of degree*

$$\deg(r) < c\ell + 2m\ell(\ell-1) + mq$$

which has a zero of order at least ℓ at all points $x \in \mathcal{S}_a$ (recall $c = \frac{1}{2}(q-1)$).

We will look, by the method of indeterminate coefficients, for a polynomial r of the special form

$$r = f^\ell \sum_{0 \leqslant j < J} (r_j + s_j g) X^{jq} \tag{11.23}$$

for some polynomials r_j, $s_j \in \mathbb{F}[X]$, to be constructed, each of which has degree bounded by $c - m$. Hence such a polynomial r has degree bounded by

$$\deg(r) \leqslant \ell m + c - m + cm + Jq \leqslant (J+m)q. \tag{11.24}$$

The next lemma is crucial to ensure that $r \neq 0$. This is where we need the assumption (11.18).

LEMMA 11.19. *We have $r = 0 \in \mathbb{F}[X]$ if and only if $r_j = s_j = 0 \in \mathbb{F}[X]$ for all j.*

PROOF. We can assume (by a shift $X \mapsto X + a$ if necessary) that $f(0) \neq 0$. Suppose that $r = 0$ but not all r_j, s_j are zero; let k be the smallest index for which one of r_k, s_k is non-zero. Dividing by $f^\ell X^{kq}$, we get from (11.22) the identity

$$\sum_{k \leqslant j < J} (r_j + s_j g) X^{(j-k)q} = 0$$

which we write in the form $h_0 + h_1 g = 0$, where

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

We square this equation, then multiply both sides by f, getting

$$h_0^2 f = h_1^2 f^q.$$

Since $f \in \mathbb{F}[X]$, we have

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

hence

$$r_k^2 f \equiv s_k^2 f(0) \pmod{X^q}.$$

However, the degree s of the polynomials in this congruence are bounded by $2\deg(r_k) + m \leqslant 2(c-m) + m < q$, and $2\deg(s_k) < 2(c-m) < q$, respectively. So there must be equality $r_k^2 f = s_k^2 f(0)$, which contradicts the assumption (11.18) that f is not a square in $\bar{\mathbb{F}}[X]$. □

We now evaluate the Hasse derivatives of r.

LEMMA 11.20. *Let $k \leqslant \ell$. Then there exist polynomials $r_j^{(k)}$, $s_j^{(k)}$ each one of degree $\leqslant c - m + k(m-1)$ such that*

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

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

$$h = f^\ell \sum_{0 \leqslant j \leqslant J} (r_j + s_j f^c) Y^{jq}.$$

Hence by Lemma 11.17, we have

$$E^k r = (E_X^k h)(X, X^q) = \sum_{0 \leqslant j < J} (E^k(f^\ell r_j) + E^k(f^{\ell+c} s_j)) X^{jq}.$$

By (2) of Corollary 11.15 there exist polynomials $r_j^{(k)}$ and $s_j^{(k)}$ satisfying $E^k(f^\ell r_j) = f^{\ell-k} r_j^{(k)}$ and $E^k(f^{\ell+c} s_j) = f^{\ell-k+c} s_j^{(k)}$ with $\deg(r_j^{(k)}) \leqslant \deg(r_j) + k\deg(f) - k \leqslant c - m + k(m-1)$ and $\deg(s_j^{(k)}) \leqslant c - m + k(m-1)$. This is the desired result. □

Recall that we wish r to have zeros of order $\geqslant \ell$ at points in $\mathcal{S}_a$ (see (11.20)). If $f(x) = 0$, clearly this is the case. So let $x \in \mathcal{S}_a$, with $f(x) \neq 0$. Applying Lemma

11.21, we evaluate $E^k r$ at a point $x \in \mathcal{S}_a$, using $g(x) = a$, and most importantly $x^q = x$:

$$\begin{aligned} E^k r(x) &= f(x)^{\ell-k} \sum_{0 \leqslant j < J} (r_j^{(k)}(x) + a s_j^{(k)}(x)) x^j \\ &= f(x)^{\ell-k} \sigma^{(k)}(x) \end{aligned}$$

where $\sigma^{(k)} \in \mathbb{F}[X]$ is the polynomial

$$\sigma^{(k)} = \sum_{0 \leqslant j < J} (r_j^{(k)} + a s_j^{(k)}) X^j.$$

We can now prove Proposition 11.18: if $\sigma^{(k)} = 0$ for all $k < \ell$, Lemma 11.16 shows that r has a zero of order $\geqslant \ell$ at all points in $\mathcal{S}_a$. The system of equations

$$\sigma^{(k)} = 0, \text{ for all } k < \ell \tag{11.25}$$

is a homogeneous system of linear equations, the unknowns being the coefficients of the polynomials r_j, s_j, the equations corresponding to the coefficients of the $\sigma^{(k)}$. We observe that

$$\deg(\sigma^{(k)}) < c - m + k(m-1) + J,$$

so the number of equations does not exceed $B = \ell(c - m + J) + \frac{1}{2}\ell(\ell-1)(m-1)$ while, on the other hand, the number of coefficients of the r_j and s_j is at least $A = 2(c-m)J$. By choosing J large enough, we can make $A > B$. Then the system (11.25) has a non-trivial solution, and by Lemma 11.19 this produces $r \neq 0$ such that r has zeros of order $\geqslant \ell$ at all points $x \in \mathcal{S}_a$. Taking

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

one can check that $A > B$ (recall that $2c = q - 1$ and $8\ell \leqslant q$). The degree of r is bounded by (11.24), which gives Proposition 11.18.

We now prove Stepanov's Theorem 11.13. First, let a be arbitrary, and apply Proposition 11.18. Since the auxiliary polynomial r is non-zero and vanishes to order $\geqslant \ell$ at points in $\mathcal{S}_a$, we have $\ell|\mathcal{S}_a| \leqslant \deg(r) \leqslant c\ell + 2m\ell(\ell-1) + mq$ so $|\mathcal{S}_a| \leqslant c + 2m(\ell-1) + mq\ell^{-1}$. We choose $\ell = 1 + [\sqrt{q}/2]$, which gives the bound

$$|\mathcal{S}_a| < c + 4m\sqrt{q}. \tag{11.26}$$

To prove Theorem 11.13, take first $a = 1$ getting

$$N_0 + N_1 = |\mathcal{S}_a| < \frac{q}{2} + 4m\sqrt{q}$$

hence the upper bound

$$N = N_0 + 2N_1 < 2(N_0 + N_1) < q + 8m\sqrt{q}. \tag{11.27}$$

To get a lower bound, by the factorization $X^q - X = X(X^c - 1)(X^c + 1)$ we have

$$f(x)(g(x) - 1)(g(x) + 1) = 0$$

for all $x \in \mathbb{F}$, hence $N_0 + N_1 + N_2 = q$ where $N_2 = |\{x \in \mathbb{F} \mid g(x) = -1\}|$. By (11.26) applied to $\mathcal{S}_{-1}$, we have

$$N_0 + N_2 = |\mathcal{S}_{-1}| < \frac{q}{2} + 4m\sqrt{q}$$

hence

$$N_1 = q - N_0 - N_2 > \frac{q}{2} - 4m\sqrt{q},$$

and finally,

$$N = N_0 + 2N_1 \geqslant 2N_1 > q - 8m\sqrt{q}. \tag{11.28}$$

Clearly (11.27) and (11.28) prove Theorem 11.13.

11.7. Proof of Weil's bound for Kloosterman sums.

Let $\mathbb{F}$ be a finite field with q elements, of characteristic $p \neq 2$. Let ψ be any fixed non-trivial additive character of $\mathbb{F}$. For any additive character φ there exists a unique $a \in \mathbb{F}$ such that $\varphi = \psi_a$, hence any Kloosterman sum $S(\psi, \varphi)$ is of the form

$$S(\psi_a, \psi_b) = -\sum_{x \in \mathbb{F}^*} \psi(ax + bx^{-1})$$

for some a, $b \in \mathbb{F}$. We consider a and b as fixed and write $g = aX + bX^{-1}$. We will prove Weil's bound (11.15) by relating the average of the Kloosterman sums $S(\psi_a, \psi_b)$ over ψ to the number of points on an hyperelliptic curve, where the contribution of the trivial character $\psi_0 = 1$ will be the main term.

LEMMA 11.21. *For any $n \geqslant 1$ and any $x \in \mathbb{F}_n$, we have*

$$|\{x \in \mathbb{F}_n \mid y^q - y = x\}| = \sum_{\psi} \psi(\mathrm{Tr}(x)) \tag{11.29}$$

where the sum ranges over all additive characters of $\mathbb{F}$ and Tr is the trace $\mathbb{F}_n \to \mathbb{F}$.

PROOF. If $\mathrm{Tr}\,(x) = 0$, then the equation $y^q - y = x$ has q solutions exactly, as recalled in Section 11.2, and in this case we have $\psi(\mathrm{Tr}\,(x)) = 1$ for all ψ, hence the right side of (11.29) is also equal to q. On the other hand, if $\mathrm{Tr}\,(x) \neq 0$, the equation $y^q - y = x$ has no solution, and the character sum is zero by orthogonality. □

From this lemma we deduce that

$$\begin{aligned} -\sum_{\psi} S_n(\psi_a, \psi_b) &= \sum_{\psi} \sum_{x \in \mathbb{F}_n^*} \psi(\mathrm{Tr}\, g(x)) \\ &= |\{(x, y) \in \mathbb{F}_n^* \times \mathbb{F}_n \mid y^q - y = g(x)\}| = N_n, \text{ say,} \end{aligned} \tag{11.30}$$

for $n \geqslant 1$. If $\psi = \psi_0$, the trivial character, we have

$$S(\psi_a, \psi_b) = S(\psi_0, \psi_0) = 1 - q^n.$$

For $\psi \neq \psi_0$, let α_ψ, β_ψ be the "roots" of the Kloosterman sum $S(\psi_a, \psi_b)$, so by Theorem 11.8 we have $\alpha_\psi \beta_\psi = q$ and

$$S_n(\psi_a, \psi_b) = \alpha_\psi^n + \beta_\psi^n,$$

for all $n \geqslant 1$.

We can therefore write

$$N_n = q^n - 1 - \sum_{\psi \neq \psi_0} (\alpha_\psi^n + \beta_\psi^n).$$

The equation $y^q - y = g(x)$ does not obviously describe a curve, since g is not a polynomial, but multiplying by x it is equivalent with

$$C_{a,b} \,:\, ax^2 - (y^q - y)x + b = 0$$

(note that $x = 0$ is not possible since $b \neq 0$). Because $p \neq 2$, the number of solutions is equal to the number of solutions of the discriminant equation of this quadratic equation

$$D_{a,b} \,:\, (y^q - y)^2 - 4ab = v^2,$$

i.e., $N_n = |D_{a,b}(\mathbb{F}_n)|$. This is of the form (11.17) with $\deg(f) = 2q$, and because $4ab \neq 0$ it satisfies (11.18). Hence by Theorem 11.13 we have

$$|N_n - q^n| < 16q^{1+n/2}$$

if n is large enough, so that $q^n > 16q$.

By (11.30) we get a sharp estimate for the roots α_ψ, β_ψ, on average

$$\frac{1}{q}\Big|\sum_{\psi\neq\psi_0}(\alpha_\psi^n + \beta_\psi^n)\Big| \leqslant 16q^{n/2} \tag{11.31}$$

for n large enough. The following simple lemma shows that the individual roots must be of modulus $\leqslant \sqrt{q}$:

LEMMA 11.22. *Let ω_1, ..., ω_r be complex numbers, A, B positive real numbers and assume that*

$$\Big|\sum_{j=1}^{r}\omega_j^n\Big| \leqslant AB^n$$

holds for all integers n large enough. Then $|\omega_j| \leqslant B$ for all j.

PROOF. One can do this by hand (using Dirichlet's box principle), but a nice trick gives the result immediately: consider the complex power series

$$f(z) = \sum_{n\geqslant 1}\Big(\sum_j \omega_j^n\Big)z^n = \sum_j \frac{1}{1-\omega_j z}.$$

The hypothesis implies that f converges absolutely in the disc $|z| < B^{-1}$, hence f is analytic in this region. In particular, it has no poles there, which means that we must have $|\omega_j|^{-1} \geqslant B^{-1}$ for all j. □

From this lemma applied with $A = 16q$, $B = \sqrt{q}$, we deduce the upper bounds $|\alpha_\psi| \leqslant \sqrt{q}$, $|\beta_\psi| \leqslant \sqrt{q}$ for all $\psi \neq \psi_0$. Since $\alpha_\psi\beta_\psi = q$, we have in fact $|\alpha_\psi| = |\beta_\psi| = \sqrt{q}$, and so Theorem 11.11 is proved.

REMARKS. (1) We see here twice how crucial the introduction of the companion sums K_n is: first because the curve $D_{a,b}$ has very high degree, so Stepanov's bound $|N-q| < 8m\sqrt{q}$ is trivial when applied to $\mathbb{F}$ itself, and secondly because only by the consideration of all extension fields can we determine the exact order of magnitude of the roots, and obtain Weil's bound $S(\psi,\varphi) \leqslant 2\sqrt{q}$ with the sharp constant 2.

(2) The constant 2 is optimal in Weil's bound for fixed a, b and q. Indeed we have

$$\sum_{n\geqslant 1} S_n(\psi_a,\psi_b)z^n = \frac{1}{1-\alpha_\psi z} + \frac{1}{1-\beta_\psi z}.$$

This is a non-zero rational function with poles on the circle $|z| = 1/\sqrt{q}$, hence this is its radius of convergence. Therefore

$$\limsup_{n\to+\infty} |S_n(\psi_a, \psi_b)|^{-1/n} = \frac{1}{\sqrt{q}}.$$

This means that for any $\varepsilon > 0$ there exist infinitely many n such that

$$|S_n(\psi_a, \psi_b)| \geqslant (2-\varepsilon)q^{n/2}.$$

It is conjectured (this follows from the Sato-Tate Conjecture for the angles of Kloosterman sums described in Chapter 21) that the Weil bound is also optimal when a, b are fixed, $n = 1$, and $q = p \to +\infty$. However, this remains very much open. See the remark at the end of the introduction to Section 11.8 for the case of elliptic curves.

(3) Using the extension of Stepanov's method to curves of the type $y^d = f(x)$ and an analysis of the corresponding zeta function, one can prove the following estimate for complete character sums:

THEOREM 11.23. *Let $\mathbb{F}$ be a finite field with q elements and let χ be a non-trivial multiplicative character of $\mathbb{F}^\star$ of order $d > 1$. Suppose $f \in \mathbb{F}[X]$ has m distinct roots and f is not a d-th power. Then for $n \geqslant 1$ we have*

$$\Big|\sum_{x\in\mathbb{F}_n} \chi(N(f(x)))\Big| \leqslant (m-1)q^{n/2}.$$

This is Theorem 2C', p. 43, of [**Sch**]. In particular, we get the following corollary which will be used in proving the Burgess bound for short character sums (Theorem 12.6).

COROLLARY 11.24. *Let $\chi \pmod p$ be a non-principal multiplicative character. If one of the classes $b_v \pmod p$, $v = 1, \dots, 2r$ is different from the remaining ones then*

$$\Big|\sum_{x \pmod p} \chi((x+b_1)\dots(x+b_r))\bar{\chi}((x+b_{r+1})\dots(x+b_{2r}))\Big| \leqslant 2rp^{\frac{1}{2}}.$$

PROOF. Observe that

$$\chi((x+b_1)\dots(x+b_r))\bar{\chi}((x+b_{r+1})\dots(x+b_{2r}) = \chi(f(x))$$

with

$$f(x) = \prod_{1\leqslant j\leqslant r}(x+b_j)\prod_{r+1\leqslant j\leqslant 2r}(x+b_j)^{p-2}.$$

From the assumption, one of the b_i is a root of f of order either 1 or $p-2$, which is coprime with the order $d \mid (p-1)$ of χ, so we can apply Theorem 11.23. □

11.8. The Riemann Hypothesis for elliptic curves over finite fields.

A particularly important case of the Riemann Hypothesis is that of elliptic curves. Historically this was first established by Hasse using global methods. In the notation of Section 11.6, this means that we consider curves C_f with $\deg f = 3$, so the equation is of the form

$$C \,:\, y^2 = x^3 + a_2x^2 + a_4x + a_6 \tag{11.32}$$

(the numbering reflects the traditional notation for elliptic curves, see Section 14.4). In contrast with that section, we emphasize that we are considering the affine curve, without the point at infinity. The cubic polynomial $f(x) = x^3 + a_2x^2 + a_4x + a_6$ cannot be a square, so this curve satisfies the assumption (11.18). Moreover, we assume that f does not have a double root; this means that the curve C is smooth (see Section 11.9), and it is a necessary condition for what follows.

In this case, Theorem 11.13 implies that for $q > 36$ the number $N = |C(\mathbb{F})|$ satisfies

$$|N - q| < 24\sqrt{q}.$$

In Section 11.10 we will prove, as before for Kloosterman sums, the rationality and the functional equation of the corresponding zeta function, from which we will deduce:

THEOREM 11.25. *Let C be an elliptic curve over $\mathbb{F}$ given by*

$$C \,:\, y^2 = x^3 + a_2x^2 + a_4x + a_6$$

with $a_i \in \mathbb{F}$. Then for all $n \geqslant 1$ we have

$$\big||C(\mathbb{F}_n)| - q^n\big| \leqslant 2q^{n/2}. \tag{11.33}$$

REMARKS. Theorem 11.25 is optimal. Indeed letting $n \to +\infty$, this follows as for Kloosterman sums from Lemma 11.22. However, it is also true in the horizontal sense as the following example shows: let $E/\mathbb{Q}$ be the elliptic curve with equation

$$E \,:\, y^2 = x^3 - x$$

which has complex multiplication by $\mathbb{Z}[i]$. As before we consider the affine points, not the projective ones. The discriminant of E is 64 so E can be reduced modulo p to an elliptic curve over $\mathbb{Z}/p\mathbb{Z}$ for any odd prime p. One shows (for instance by relating E to the curve $y^2 = x^4 + 4$ by changing $(x, y) \mapsto (yx^{-1}, 2x - y^2x^{-2})$ for $(x, y) \neq (0, 0)$, see e.g. [**I4**] or [**IR**]) that $|E(\mathbb{Z}/p\mathbb{Z})| = p$ if $p \equiv 3 \pmod 4$ and $|E(\mathbb{Z}/p\mathbb{Z})| = p - 2a_p$ if $p \equiv 1 \pmod 4$, where

$$p = a_p^2 + b_p^2$$

with $\pi = a_p + ib_p \equiv 1 \pmod{2(1+i)}$ (this congruence determines π up to conjugation). For any $\varepsilon > 0$, Theorem 5.36 (generalized slightly to add the congruence condition) shows that there exist infinitely many Gaussian primes $\pi \equiv 1 \pmod{2(1+i)}$ such that $|\arg \pi| < \varepsilon$. Hence $|\mathrm{Im}\,(\pi)| \leqslant \varepsilon|\pi|$ and

$$|p - |E(\mathbb{Z}/p\mathbb{Z})|| = 2|a_p| \geqslant 2(1 - \varepsilon^2)\sqrt{p}$$

for infinitely many p.

Theorem 11.25 will be proved in Section 11.10 after some geometric and algebraic preliminaries. This goes a bit further away from the heart of analytic number theory, yet we include full details because the Hasse bound is also important as being the simplest case of the very important Deligne bound for Fourier coefficients of modular forms. The reader will also certainly appreciate the elegance and beauty of the geometry involved.

11.9. Geometry of elliptic curves.

In explaining the special geometric features of elliptic curves, we may as well consider a more general case. So let k be an arbitrary field, $\bar{k}$ an algebraic closure, and let C be the curve given by the equation

$$C : y^2 = f(x), \text{ with } f = X^3 + a_2X^2 + a_4X + a_6 \in k[X],$$

identified with the set of solutions $(x, y) \in \bar{k}^2$. We assume as before that f does not have a double root. If k'/k is any extension, we let $C(k')$ be the set of solutions in $(k')^2$.

The geometry of the elliptic curve becomes much clearer if we work with the projective version of the curve C, namely the curve E in the projective plane given, in homogeneous coordinates $(x : y : z)$, by the equation

$$E : y^2z = x^3 + a_2x^2z + a_4xz^2 + a_6z^3.$$

Putting $z = 1$ gives back C; on the other hand, "at infinity", we are only adding one point: taking $z = 0$ yields $x = 0$, and all elements $(0 : y : 0)$ (with $y \neq 0$) correspond to a single point $\infty = (0 : 1 : 0)$ in the projective plane. Notice that this point ∞ is rational over the base field k, so that for all extensions k'/k we have

$$E(k') = C(k') \cup \{\infty\}.$$

The main property of the curve E that we will use is the beautiful fact that its points form an abelian group, with identity element ∞. Throughout, p denotes points on E, not the characteristic of the field k. The group law (denoted by $+$) is described by the geometric condition that for any three (distinct) points p_1, p_2 and p_3 in E, we have $p_1 + p_2 + p_3 = 0$ if and only if the three points are collinear (in the projective plane), and the opposite of a point $(x : y : z)$ is the point $(x : -y : z)$ (symmetry with respect to the x-axis). This way one can construct the sum of any two distinct points, by computing the equation of the line joining them, and taking the opposite (in the sense above) of the third intersection point with the curve. That there are exactly three intersection points follows immediately from the fact that the polynomial f is of degree 3. In addition, to compute the double $p + p$ of a point p, the same construction is done with the tangent line at p; the condition that f has no double root ensures that this tangent line always exists.

Also, because $f \in k[X]$, it follows easily that the k'-rational points $E(k')$, for any extension k'/k, form a subgroup of $E(\bar{k})$.

We do not prove those facts here; completely elementary proofs, by computing explicitly the coordinates of the sum $p_1 + p_2$ of two points according to the recipe above and checking the abelian group axioms (associativity is the only difficulty), are fairly straightforward (see for instance [**IR**], ch. 18, 19).

We now introduce some further geometric objects related to E or, more generally, to any smooth, projective, algebraic curve[1]. Thus consider again a more general case: let $\bar{k}$ be an algebraically closed field, and let E be a plane algebraic curve over $\bar{k}$, i.e. given by an equation

$$f(x, y, z) = 0$$

for some homogeneous $f \in \bar{k}[X, Y, Z]$. We identify E with the set of points in the projective plane. We assume that E is smooth, which means here that for any point $p = (x : y : z)$ of E, not all partial derivatives $\partial f/\partial X(p)$, $\partial f/\partial Y(p)$, $\partial f/\partial Z(p)$, are zero. In this case the line with equation

$$\frac{\partial f}{\partial X}(p)(X - x) + \frac{\partial f}{\partial Y}(p)(Y - y) + \frac{\partial f}{\partial Z}(p)(Z - z) = 0$$

is well-defined and is the tangent line to E at p. For elliptic curves $y^2 = f(x)$, this smoothness condition is equivalent to the fact that the polynomial f has no double roots, by a simple calculation.

Let C be the affine curve corresponding to E given by

$$C \,:\, f(x, y, 1) = 0$$

in $\bar{k}^2$. Let $g(X, Y) = f(X, Y, 1) \in \bar{k}[X, Y]$. We define $\bar{k}[C] = \bar{k}[X, Y]/(g)$. Elements of $\bar{k}[C]$ can be interpreted as functions on C. We assume that (g) is a prime ideal (this is easily checked in the case of elliptic curves) so that $\bar{k}[C]$ is an integral domain, and we let $\bar{k}(C)$ or $\bar{k}(E)$ be its quotient field, called the function field of C or of E. It is a finite extension of the field $\bar{k}(X)$ of rational functions over $\bar{k}$ (for elliptic curves $y^2 = f(x)$, it is a quadratic extension $\bar{k}(X)(\sqrt{f})$). We interpret elements of the function field as rational functions on E, so given a point $p \in E$ and an element $\varphi \in \bar{k}(E)$, either φ has a pole at p or $\varphi(p) \in \bar{k}$ is defined.

Now the important point is that because E is smooth it is possible to define the order of φ at p for every p in E and every non-zero rational function $\varphi \in \bar{k}(E)^\star$. As expected, this order behaves like its analogue for holomorphic functions theory or rational functions. Precisely, for every $p \in R$, there is a discrete valuation

$$\mathrm{ord}_p \,:\, \bar{k}(E)^\times \to \mathbb{Z},$$

which gives the order of the zero (if $\geqslant 0$) or pole (if < 0) of a rational function at p. As a discrete valuation, it satisfies

$$\begin{aligned} \mathrm{ord}_p(c) &= 0, \text{ for } c \in \bar{k}^\star, \\ \mathrm{ord}_p(\varphi\psi) &= \mathrm{ord}_p(\varphi) + \mathrm{ord}_p(\psi), \\ \mathrm{ord}_p(\varphi + \psi) &\geqslant \min(\mathrm{ord}_p(\varphi), \mathrm{ord}_p(\psi)). \end{aligned}$$

We sketch a proof (see also [**Sil**], Prop. II.1.1): consider the ring $\mathcal{O}_p = \{\varphi \in \bar{k}(E) \mid \varphi \text{ is defined at } p\}$. This is a noetherian local domain with maximal ideal $\mathfrak{m}_p = \{\varphi \in \mathcal{O}_p \mid \varphi(p) = 0\}$ and residue field $\mathcal{O}_p/\mathfrak{m}_p \simeq \bar{k}$ (by evaluation at p). Because E is smooth at p (there is a tangent line), the $\bar{k}$-vector space $\mathfrak{m}_p/\mathfrak{m}_p^2$ is of dimension 1 (because the curve is in the plane, it is of dimension $\leqslant 2$; the equation of the tangent line gives a relation, and it is easy to see that $\mathfrak{m}_p/\mathfrak{m}_p^2 \neq 0$). By Nakayama's Lemma (see for instance [**AM**], p. 21), it follows that $\mathfrak{m}_p$ is a principal

[1]The reader can without damage assume that we are just dealing with the elliptic curves described before, with $k = \bar{\mathbb{F}}$. In this case every incomplete assertion can be checked by hand.

ideal. Let π be a generator; then $\mathfrak{m}_p^d$ is generated by π^d for any $d \geqslant 1$. The order ord_p can be defined for φ in $\mathcal{O}_p$, $\varphi \neq 0$, by

$$\mathrm{ord}_p(\varphi) = \max\{d \geqslant 0 \mid \varphi \in \mathfrak{m}_p^d\}$$

and extended to a homomorphism $\bar{k}(E)^\star \to \mathbb{Z}$. The properties above are then quite easy to check.

For elliptic curves $y^2 = f(x)$, one can easily see that if $p \neq \infty$, and $p = (x, y)$ with $y \neq 0$, it is possible to take $\pi = X - x$. For $p = \infty$, one can take $\pi = X/Y$, and one finds that $\mathrm{ord}_\infty(x) = -2$, $\mathrm{ord}_\infty(y) = -3$.

Every non-zero element $\varphi \in \bar{k}(E)$ has finitely many zeros and poles, and to package them conveniently we define a divisor on E to be a formal finite linear combination with coefficients in $\mathbb{Z}$ of symbols $[p]$, one for each point $p \in E$. Divisors form a free abelian group $\mathrm{Div}(E)$. Two homomorphisms are important. One associates to a non-zero rational function φ the divisor (denoted (φ)) of its zeros and poles:

$$\begin{cases} \bar{k}(E)^\times \to \mathrm{Div}(E) \\ \varphi \mapsto (\varphi) = \displaystyle\sum_{p \in E} \mathrm{ord}_p(\varphi) \end{cases}$$

and the second gives the degree of a divisor:

$$\deg \begin{cases} \mathrm{Div}(E) \to \mathbb{Z}, \\ [p] \mapsto 1. \end{cases}$$

As suggested by the notation, divisors of the type (φ) are called principal divisors.

Ordinary rational functions have as many zeros as poles, with multiplicity, and the same holds for $\varphi \in \bar{k}(E)^\star$: this means that for all $\varphi \in \bar{k}(E)^\times$, we have $\deg((\varphi)) = 0$ (see for instance [**Sil**], II-3). For elliptic curves $y^2 = f(x)$, an easy proof can be derived by observing that $\bar{k}(E)$ is a quadratic extension of $\bar{k}(X)$. The non-trivial element in the Galois group is $\varphi \mapsto \bar{\varphi}$ defined by $\bar{\varphi}(p) = \varphi(-p)$. It is clear that $\mathrm{ord}_p(\varphi) = \mathrm{ord}_{-p}(\bar{\varphi})$, hence $\deg(\varphi) = \deg(\bar{\varphi})$. Now $\varphi\bar{\varphi}$ is in $\bar{k}(X)$. One can check the following compatibility: if $\psi \in \bar{k}(X)$, with divisor $(\psi)_1 = \sum n_i x_i$ (as an ordinary rational function), then its divisor as an element of $\bar{k}(E)$ is $(\psi) = \sum n_i([p_i] + [-p_i])$, where p_i is any point of E with x-coordinate x_i. In particular, $0 = \deg((\psi)_1) = 2\deg((\psi))$. Applying this to $\varphi\bar{\varphi}$ gives

$$0 = \deg(\varphi\bar{\varphi}) = \deg(\varphi) + \deg(\bar{\varphi}) = 2\deg(\varphi).$$

Definition. 1. Two divisors D_1 and D_2 such that $D_1 - D_2 = (\varphi)$ for some $\varphi \in \bar{k}(E)^\times$ are called linearly equivalent. This is denoted $D_1 \sim D_2$, and is an equivalence relation on $\mathrm{Div}(E)$.

2. The group $\mathrm{Div}(E)$ carries a partial ordering, compatible with the group structure, defined by $D \geqslant 0$ if and only if all the coefficients in the formal sum giving D are $\geqslant 0$. Such divisors are called effective divisors.

3. For a divisor $D \in \mathrm{Div}(E)$, let

$$L(D) = \{0\} \cup \{\varphi \in \bar{k}(E) \mid (\varphi) + D \geqslant 0\};$$

this is a $\bar{k}$-vector space. Let $\ell(D) = \dim L(D)$, an integer or $+\infty$.

If $D = n_1[p_1] + \ldots + n_k[p_k] - m_1[q_1] - \ldots - m_j[q_j]$ with $n_i, m_i \geqslant 0$, then $\varphi \in L(D)$, $\varphi \neq 0$, means simply that φ has

(1) Poles of order at most n_i at p_i, $1 \leqslant i \leqslant k$;
(2) Zeros of order at least m_i at q_i, $1 \leqslant i \leqslant j$.

It follows immediately that if $D_2 \leqslant D_1$, then $L(D_2) \subset L(D_1)$. Also, if $D_1 \sim D_2$, then writing $D_1 = D_2 + (\psi)$, the map $\varphi \mapsto \psi\varphi$ induces an isomorphism $L(D_1) \to L(D_2)$, in particular, $\ell(D_1) = \ell(D_2)$ only depends on the linear equivalence class of the divisor.

The following interpretation is also clear: there is a bijection

$$\begin{cases} \mathbf{P}(L(D)) \to \{\text{Effective divisors linearly equivalent to } D\}, \\ \varphi \mapsto (\varphi) + D \end{cases} \tag{11.34}$$

between the projective space $\mathbf{P}(L(D))$ of $L(D)$ and the effective divisors linearly equivalent to D (by definition of $L(D)$, the map has image in the set of effective divisors). This also requires the important fact that

$$L(0) = \bar{k}, \quad \ell(0) = 1. \tag{11.35}$$

In other words, an everywhere defined rational function on E is constant: this is obvious for elliptic curves, since regularity on C forces such a φ to be a polynomial $g(X) + Yh(X)$, and regularity at ∞ then forces $h = 0$, $g \in \bar{k}$.

We first notice the following simple lemma:

LEMMA 11.26. *Let D be a divisor on E such that $\ell(D) > 0$. Then either* $\deg(D) > 0$ *or* $D \sim 0$.

PROOF. If $\ell(D) > 0$, there is a non-zero element $\varphi \in L(D)$, so that $(\varphi)+D \geqslant 0$. Taking the degree we find that $\deg(D) \geqslant 0$. So we need only show now that if $\deg(D) = 0$, then $D \sim 0$. But $(\varphi) + D$ is then an effective divisor of degree 0; clearly it must be $= 0$, so $D = -(\varphi) = (\varphi^{-1}) \sim 0$. □

To prove the rationality of the local zeta function of elliptic curves, we will need to know the value of $\ell(D)$ for effective divisors. This is computed by the Riemann-Roch Theorem.

THEOREM 11.27. *Let E be an elliptic curve over an algebraically closed field $\bar{k}$. For any divisor D on E, $\ell(D)$ is finite and we have the formula*

$$\ell(D) - \ell(-D) = \deg D. \tag{11.36}$$

Equivalently, by Lemma 11.26, we can compute $\ell(D)$ for any D by:
1. If $\deg(D) \geqslant 0$, *and* $D \not\sim 0$, *then* $\ell(D) = \deg(D)$.
2. If $D \sim 0$, *then* $\ell(D) = 1$.
3. If $\deg(D) < 0$, *then* $\ell(D) = 0$.

This is the Riemann-Roch theorem specialized for elliptic curves. See the remark below for the general case.

The simple proof for the Riemann-Roch theorem hinges on the remarkable interaction of the group structure on E with the divisor group. Indeed, consider the map

$$\sigma : \mathrm{Div}(E) \to E$$

defined by $\sigma(n_1[p_1] + ... + n_k[p_k]) = n_1p_1 + ... + n_kp_k$, the $+$ on the right side corresponding to the group law on E.

PROPOSITION 11.28. *Let D be a divisor on E. Then we have*

$$D \sim [\sigma(D)] + (\deg(D) - 1)[\infty].$$

PROOF. In essence this "is" the group law itself: by induction, we need only consider $D = [p] + [q]$ and $D = [p] - [q]$ for some points p and q. If p or q is the origin ∞, the result is obvious.

So consider first the case of $D = [p] + [q]$ with $p,\ q \neq \infty$ and $p \neq q$. Then the equation $aX + bY + c = 0$ of the line joining p and q defines an element $\varphi = aX + bY + c \in \bar{k}(E)$, which by definition of the group law satisfies

$$(\varphi) = [p] + [q] + [r] - 3[\infty]$$

where $-r = p + q = \sigma(D)$. Since $(\varphi) \sim 0$ we get

$$D = [p] + [q] \sim (\varphi) - [\sigma(D)] + 3[\infty] \sim -[\sigma(D)] + 3[\infty].$$

But similarly for any point p in E, the equation of the line joining p and $-p$ gives a function with divisor $[p] + [-p] - 2[\infty]$ so that for any p

$$-[p] + [\infty] \sim [-p] - [\infty] \tag{11.37}$$

and hence $D \sim [\sigma(D)] + [\infty]$, as desired.

If $p = q$, the equation of the tangent line gives $2[p] + [2p] - 2[\infty] \sim 0$, and the result again follows. Finally if $D = [p] - [q]$, use (11.37) to reduce to the previous case:

$$[p] - [q] = [p] + (-[q] + [\infty]) - [\infty] \sim [p] + [-q] - 2[\infty] \sim [p + q] - [\infty].$$

□

PROOF OF THE RIEMANN-ROCH THEOREM. Notice that (11.36) for D or $-D$ are equivalent. By Proposition 11.28, we have

$$\ell(D) = \ell([\sigma(D)] + (\deg D - 1)[\infty]),$$
$$\ell(-D) = \ell([-\sigma(D)] + (1 - \deg D)[\infty]).$$

One of the two divisors on the right is effective, so we can assume that D is effective and of the form $D = [p] + n[\infty]$ with $P \in E$ and $n \geqslant 0$.

If $p = \infty$, then $D = m[\infty]$ with $m \geqslant 1$. We must prove $\ell(D) = m$. Any element φ of $L(m[\infty])$ has no poles on C, hence is a polynomial $\varphi \in \bar{k}[X, Y]$, which satisfies $\operatorname{ord}_\infty(\varphi) \geqslant -m$. Since $\operatorname{ord}_\infty(X) = -2$ and $\operatorname{ord}_\infty(Y) = -3$, we have

$$\operatorname{ord}_\infty(\varphi) = \max(-2\deg(g), -3 - 2\deg(h)) \text{ for } \varphi = g(X) + Yh(X).$$

Let $V = \{2\deg(g), 3 + 2\deg(h)\}$ where g and h are polynomials in X. One sees immediately that V is the set of all positive integers, except 1. Now using monomials X^a or YX^b as basis elements, we check that

$$\ell(m[\infty]) = |\{n \in V \mid n \leqslant m\}| = m.$$

Now we consider $D = [p] + n[\infty]$ with $n \geqslant 0$ and $p \neq \infty$. If $n = 0$, we can use the automorphism $q \mapsto q - p$ which sends p to ∞ to get an isomorphism $L([p]) \simeq L([\infty])$ which implies $\ell([p]) = 1$.

If $n \geqslant 1$, we have $D \geqslant n[\infty]$ hence $L(n[\infty]) \subset L(D)$. Thus $n \leqslant \ell(D)$. Moreover, because only a simple pole is allowed at p, we have $\ell(D) \leqslant n+1$: if π_p is a function with a simple zero at p, we have a $\bar{k}$-linear map

$$\begin{cases} L(D)/L(n[\infty]) \to \bar{k} \\ \varphi \mapsto (\pi_p \varphi)(p) \end{cases}$$

which is tautologically injective. Thus we must show that there is in $L(D)$ one more $\bar{k}$-linearly independent element not in $L(n[\infty])$.

Write $p = (x, y)$ in affine coordinates. We have the following divisors:

$$(X - x) = [p] + [-p] - 2[\infty],$$
$$(Y + y) = [-p] + [p'] + [p''] - 3[\infty]$$

for some p' and p''. Let $\varphi = (Y+y)/(X-x)$, then $(\varphi) = -[p]+[p']+[p'']-[\infty] \geqslant -D$. We claim that $p \neq p'$, p''. Indeed p' and p'', by definition, are of the form $(x_1, -y)$, $(x_2, -y)$. This can be equal to p only if $y = 0$, but for $y = 0$ by assumption there are three distinct roots of $f(x) = 0$. Hence φ has a simple pole at p, so $\varphi \notin L(n[\infty])$, and we obtain the required formula $\ell(D) = n + 1 = \deg(D)$. □

We must now consider some rationality questions. We assume that we have an elliptic curve E given by $y^2 = f(x)$ with $f \in k[X, Y]$. The preceding analysis applies to an algebraic closure $\bar{k}$ of k. There is a natural action of the Galois group G_k of k on E (on the coordinates), and on the divisors. A point or a divisor is called k-rational if it is G_k-fixed. Notice that for a divisor $D = n_1 p_1 + ... + n_j p_j$ this does not mean that the p_i are in $E(k)$. Also G_k acts on $\bar{k}(E)$ and the field fixed by G_k is $k(E)$, the fraction field of $k[X, Y]/(f)$. The divisor of a function $\varphi \in k(E)$ is obviously k-rational.

If D is defined over k, we define

$$L_k(D) = \{0\} \cup \{\varphi \in k(E) \mid (\varphi) + D \geqslant 0\}$$

and we let $\ell_k(D) = \dim_k L_k(D)$. It is clear that $\ell_k(D) \leqslant \ell(D)$.

THEOREM 11.29. *For any k-rational divisor D, we have*

$$\ell_k(D) = \ell(D).$$

In particular the Riemann-Roch formula holds with $\ell_k(D)$ instead of $\ell(D)$.

PROOF. This is a special case of the following theorem, which is a formulation of Hilbert's Theorem 90 for $GL(n)$: let k be a field, $\bar{k}$ an algebraic closure, V a $\bar{k}$-vector space with an action of G_k. Then there is a basis of V made of elements which are G_k-fixed (equivalently, let $V_k = V^{G_k}$, then $V = V_k \otimes \bar{k}$, or $\dim_k V_k = \dim_{\bar{k}} V$). For a proof, see for instance [**Sil**], Lemma II.5.8.1. □

REMARK. This theory adapts in the following way to more general algebraic curves (see for instance [**Ha**], IV): if E is smooth and projective over $\bar{k}$, one can define a certain divisor class K (called the canonical class, and related to differentials on E). It has degree $\deg(K) = 2g - 2$ for some integer $g \geqslant 0$, called the genus of E, and the Riemann-Roch Theorem takes the form

$$\ell(D) - \ell(K - D) = \deg(D) + 1 - g. \tag{11.38}$$

Elliptic curves correspond to $g = 1$; in this case the canonical class is trivial, and this reduces to Theorem 11.27. The case $g = 0$ corresponds to the projective line, and is also very easy. The proof of (11.38) is much more involved than the one for elliptic curves, since there is no group law on the curve which would help.

11.10. The local zeta function of elliptic curves.

Let C be an elliptic curve over $\mathbb{F}$ given by (11.32). It is more convenient to use here the corresponding projective curve E, as described in Section 11.9. The zeta function of E is defined as the formal power series

$$Z(E) = \exp\Big(\sum_{n\geqslant 1} \frac{|E(\mathbb{F}_n)|}{n} T^n\Big). \tag{11.39}$$

We first relate $Z(E)$ to points on the curve, by giving its Euler product (compare Lemma 11.7). To do this we introduce some terminology, which comes from the language of schemes.

Definition. Let E be an elliptic curve over a finite field $\mathbb{F}$ with q elements. A closed point of E is the Galois orbit of a point $x_0 \in E(\overline{\mathbb{F}})$. The degree $\deg(x)$ of a closed point x is the cardinality (necessarily finite) of the orbit and its norm is $Nx = q^{\deg(x)}$. The set of closed points of E is denoted $|E|$.

This notion is analogue to that of an irreducible polynomial in $\mathbb{F}[X]$ used for the zeta functions of Gauss sums and Kloosterman sums. To every closed point $x \in |E|$ is associated an $\mathbb{F}$-rational divisor which is simply the formal sum of all the elements in the orbit. The degree of this divisor is the degree of x. Moreover, it is easy to see that the group of $\mathbb{F}$-rational divisors is the free abelian group generated by the divisors associated to closed points.

Lemma 11.30. *We have the Euler product expansion*

$$Z(E) = \prod_{x\in|E|} (1 - T^{\deg(x)})^{-1}, \tag{11.40}$$

where the product is over all closed points of E.

Proof. This is very close to Lemmas 11.7 and 11.9. First by decomposing the points in $E(\mathbb{F}_n)$ in Galois orbits we obtain

$$|E(\mathbb{F}_n)| = \sum_{d|n} d \sum_{\substack{x\in|E|\\ \deg(x)=d}} 1,$$

which is the analogue of Lemma 11.1. Then we have

$$-T\frac{Z'(E)}{Z(E)} = \sum_{n\geqslant 1} |E(\mathbb{F}_n)| T^n$$

and, on the other hand, this operator applied to the right side of (11.40) yields

$$\sum_{x\in|E|} \deg(x) \sum_{n\geqslant 1} T^{n\deg(x)} = \sum_{n\geqslant 1} T^n \Big(\sum_{d|n} d \sum_{\substack{x\in|E|\\ \deg(x)=d}} 1\Big)$$

hence the result. □

Using the Riemann-Roch Theorem, we now prove the rationality and functional equation of the zeta function.

THEOREM 11.31. *The zeta function $Z(E)$ of an elliptic curve is a rational function. More precisely, it is of the form*

$$Z(E) = \frac{1 - aT + qT^2}{(1-T)(1-qT)} \tag{11.41}$$

where $a \in \mathbb{Z}$ is defined by the relations $|E(\mathbb{F})| = q+1-a$ or, in terms of C, $|C(\mathbb{F})| = q - a$. The zeta function satisfies the functional equation $Z(E, (qT)^{-1}) = Z(E, T)$.

LEMMA 11.32. *Let $d \geqslant 0$ and let $h_d(C)$ be the set of linear equivalence classes of $\mathbb{F}$-rational divisors of degree d. Then $h_d(C)$ is finite and $|h_d(C)| = |h_0(C)| \leqslant |E(\mathbb{F})|$.*

PROOF. For any rational divisor D, we have by Proposition 11.28 the equivalence $D \sim [\sigma(D)] + (\deg D - 1)[\infty]$, thus the linear equivalence class of D only depends on $\sigma(D)$. If D is $\mathbb{F}$-rational, it follows that $\sigma(D) \in E(\mathbb{F})$, and therefore the inequality $|h_d(C)| \leqslant |E(\mathbb{F})|$ holds.

Moreover, it is clear that the map $D \mapsto D + d[\infty]$ with inverse $D \mapsto D - d[\infty]$ induces a bijection between $h_d(C)$ and $h_0(C)$. □

PROOF OF THEOREM 11.31. Since $\mathbb{F}$-rational divisors on E are simply combinations with integer coefficients of divisors associated to closed points, the Euler product (11.40) gives the formal power series expression

$$Z(E) = \sum_{D \geqslant 0} T^{\deg(D)}$$

where the sum is over all effective $\mathbb{F}$-rational divisors on E.

Split the sum according to the degree d of D; for $d = 0$, the only effective divisor is $D = 0$ so

$$Z(E) = 1 + \sum_{d \geqslant 1} T^d \sum_{\substack{D \geqslant 0 \\ \deg(D) = d}} 1.$$

For each d, split further the sum over divisors of degree d in linear equivalence classes. By Lemma 11.32, there are $h_0(C)$ equivalence classes for each d. For a given class (that of D say), the contribution is the number of effective ($\mathbb{F}$-rational) divisors linearly equivalent to D. By (11.34) and Theorem 11.29, this is equal to

$$|\mathbf{P}(L(D))| = \frac{q^{\ell_{\mathbb{F}}(D)} - 1}{q-1} = \frac{q^{\ell(D)} - 1}{q-1}.$$

Since $d \geqslant 1$, the Riemann-Roch theorem implies $\ell(D) = \deg(D) = d$, so the computation of $Z(E)$ is now straightforward

$$\begin{aligned} Z(E) &= 1 + \sum_{d \geqslant 1} T^d \sum_{\substack{D \geqslant 0 \\ \deg(D) = d}} 1 = 1 + \frac{h_0(C)}{q-1} \sum_{d \geqslant 1} (q^d - 1) T^d \\ &= 1 + \frac{h_0(C)}{q-1} \Big(\frac{T^d}{1-qT} - \frac{T}{1-T} \Big) = 1 + \frac{h_0(C) T}{(1-T)(1-qT)} \\ &= \frac{1 - bT + qT^2}{(1-T)(1-qT)} \end{aligned} \tag{11.42}$$

where b is defined by $h_0(C) = q + 1 - b$.

This proves the rationality, and gives the precise form, except that we need to prove that $a = b$, where $|E(\mathbb{F})| = q+1-a$, or equivalently $h_0(C) = |E(\mathbb{F})|$ (actually in Lemma 11.32, we have already shown $|h_0(C)| \leqslant |E(\mathbb{F})|$, but we do not need it any more). To obtain this equality, start from the original definition (11.39) of $Z(E)$, and compare with (11.42): the latter is seen to imply that $|E(\mathbb{F})| = q+1-b = h_0(C)$.

Finally the functional equation of $Z(E)$ is a formal consequence of (11.42). □

It is worth recording separately one of the last steps of the proof.

PROPOSITION 11.33. *Let E be an elliptic curve over a finite field $\mathbb{F}$, let D be a divisor on E. Then D is principal if and only if* $\deg(D) = 0$ *and* $\sigma(D) = 0 \in E$. *More precisely, the map $j : D \mapsto \sigma(D)$ is an isomorphism between the group of divisor classes of degree* 0 *and $E(\overline{\mathbb{F}})$.*

PROOF. A divisor D is $\mathbb{F}_n$-rational for some $n \geqslant 1$; looking at E over $\mathbb{F}_n$, it suffices to prove the isomorphism between classes of $\mathbb{F}$-rational divisors of degree 0 and $\mathbb{F}$-rational points. But j is a surjective ($j([x] - [\infty]) = x$) map between finite sets with the same cardinality ($|h_0(C)| = |E(\mathbb{F})|$). □

This is the special case of the so-called Abel-Jacobi Theorem, for an elliptic curve over a finite field. It actually holds over any field, and a generalization to all (smooth projective) curves is the content of the theory of jacobian varieties associated to curves.

To conclude the proof of Theorem 11.25, we proceed as in the case of Kloosterman sums: from (11.41), we derive

$$|E(\mathbb{F}_n)| - (q^n + 1) = \alpha^n + \beta^n$$

where $1 - aT + qT^2 = (1 - \alpha T)(1 - \beta T)$. Then Lemma 11.22, applied with the input from Stepanov's Theorem 11.13, shows that $|\alpha| \leqslant \sqrt{q}$, $|\beta| \leqslant \sqrt{q}$, and since $\alpha\beta = q$, this concludes the proof.

EXERCISE 2. Assuming the general Riemann-Roch formula (11.38), prove that for a smooth projective algebraic curve E of genus g over a finite field $\mathbb{F}$ with q elements, the zeta function

$$Z(E) = \exp\Big(\sum_{n\geqslant 1} \frac{|E(\mathbb{F}_n)|}{n} T^n\Big)$$

is a rational function of the form

$$Z(E) = \frac{P(T)}{(1 - T)(1 - qT)}$$

for some polynomial P with integral coefficients and degree $2g$.

[**Hint:** The question will arise whether there exist $\mathbb{F}$-rational divisor classes on E of degree 1 (which is obvious for elliptic curves since the point ∞ is $\mathbb{F}$-rational). The image of the degree map is $\delta\mathbb{Z}$ for some $\delta \mid (2g-2)$ (the degree of the canonical class). Using this fact, find a preliminary form of the zeta function and analyze the poles to show that actually $\delta = 1$ (see [**Mor2**], 3.3).]

11.11. Survey of further results: a cohomological primer.

The methods of Stepanov are very useful and, in certain circumstances they provide the best tools available today, especially when the genus of the curve is large compared to the cardinality of the finite field (see for instance the proof by Heath-Brown of non-trivial estimates for Heilbronn sums [**HB2**]).

However, the deepest understanding of exponential sums over finite fields and the greatest impact on classical problems of analytic number theory comes from the sophisticated concepts of algebraic number theory, especially the ℓ-adic cohomology theory as developed by Grothendieck and his collaborators, which give a very powerful and flexible framework for working with very general exponential sums.

The proof of the Riemann Hypothesis for varieties by Deligne [**De1**], and even more his far-reaching generalization [**De2**], are the basis for the extensive work of Katz, Laumon and others. It is beyond the scope of this book to discuss this theory in great detail. Let us direct the interested reader to the survey articles [**Lau**], [**K2**]. Study of the foundational basis of the ℓ-adic theory can be started in [**De3**] and continued together with applications in the books of Katz, for instance [**K3**], [**K4**].

We will limit this section to a short introduction of the basic vocabulary and we will state a few of the most fundamental results in this language. We then include examples to show that such knowledge can already be very useful even when one is not familiar with the details and background of algebraic geometry.

In Sections 11.4 and 11.5, we have shown that Gauss sums and Kloosterman sums can be related to analogues of Dirichlet characters over finite fields. The ℓ-adic cohomological formalism which we now discuss can be thought as relating exponential sums, dually, to objects which are Galois-theoretic in nature.

The exponential sums S_n defined by (11.8) can be interpreted as sums over the algebraic curve $U_{f,g}$ consisting of $\overline{\mathbb{F}}_p^*$ minus the poles of the rational functions f and g. More generally, one wishes to consider exponential sums not only over curves but over more general varieties. We will use some basic vocabulary of algebraic geometry to describe such situations, but will illustrate them in the simpler case of curves. Already the case of (11.8) and $U_{f,g}$ are quite interesting.

Let $\mathbb{F}$ be a finite field and $U/\mathbb{F}$ be a smooth algebraic variety of dimension $d \geqslant 0$ (technically, we assume as part of the smoothness assumption that U is geometrically connected, and as part of being a variety that U is quasi-projective). The simplest examples in dimension 1 are $U_{f,g}$, or smooth projective curves. In dimension $d > 1$, the most important examples are the affine d-space $\mathbb{A}^d$, with set of points $\mathbb{A}^d(\overline{\mathbb{F}}) = \overline{\mathbb{F}}^d$, and the projective d-space. The exponential sums over U will be of the type

$$S_n = \sum_{x \in U(\mathbb{F}_n)} \chi(N(f(x)))\psi(\mathrm{Tr}\,(g(x))) \tag{11.43}$$

where f and g are $\mathbb{F}$-rational functions defined on U.

To $U/\mathbb{F}$ is associated the so-called arithmetic étale fundamental group $\pi_1(U)$ which "classifies" étale coverings $V \to U$ of U, and is the analogue both of the Galois group of a field, or of the "ordinary" topological fundamental group. A morphism of algebraic varieties is étale if it is flat and unramified; if U is a curve, this means V is a curve, f is non-constant and unramified. For the simpler purposes of exponential sums, the fundamental group can be considered somewhat as a black

box in what follows, but one should keep in mind that the elements of V in $\pi_1(U)$ act as automorphisms of any étale covering $\pi : V \to U$ (i.e. $\pi(\gamma x) = \pi(x)$ for any $\gamma \in \pi_1(U)$ and $x \in V$), and that it is a functor: any map $U \to V$ between varieties induces a continuous group homomorphism $\pi_1(U) \to \pi_1(V)$. (One should fix a base-point in defining $\pi_1(U)$, but a more or less canonical choice exists, the so-called "generic point" of the scheme U.)

EXAMPLES. (1) Let U be a single point $\{x\}$ defined over $\mathbb{F}$. Then $\pi_1(U)$ is the Galois group $\mathrm{Gal}(\overline{\mathbb{F}}/\mathbb{F})$.

(2) Let $U/\mathbb{F}$ be a smooth curve, not necessarily projective. There is a an associated smooth projective curve $C/\mathbb{F}$ such that $U \subset C$ with complement a finite set T of points. If $U = \overline{\mathbb{F}}^*$ for instance, then $C = \mathbb{P}^1$ is the projective line, and $T = \{0, \infty\}$.

The fundamental group can be described concretely as follows: let $K = \mathbb{F}(U) = \mathbb{F}(C)$ be the function field of U, i.e. the field of rational functions on U or C (if $C = \mathbb{P}^1$, then $K = \mathbb{F}(t)$ is the usual field of rational fractions). We have the Galois group $G_K = \mathrm{Gal}(\bar{K}/K)$ of K. For every closed point x of C, there is the corresponding discrete valuation ord_x of K. This extends to the separable closure $\bar{K}$ of K, and gives rise to a decomposition group $D_x < G_K$ and an inertia group $I_x < D_x$ as in classical algebraic number theory, with the property that $D_x/I_x \simeq \mathrm{Gal}(\bar{\mathbb{F}}_q/\mathbb{F}_q)$, where $\mathbb{F}_q$ is the residue field of x, a finite field with $q = Nx$ elements. Then $\pi_1(U)$ "is" the quotient of G_K by the smallest closed normal containing all inertia groups I_x for x a closed point of U.

Fix a prime number $\ell \neq p$. The objects used to interpret exponential sums over U are the so-called ℓ-adic sheaves on U. In the simpler cases, those will be "lisse", in which case there is a simpler alternate Galois-theoretic description which we take as definition.

DEFINITION. Let $U/\mathbb{F}$ be a smooth variety over a finite field. A lisse ℓ-adic sheaf on U is a continuous representation $\rho : \pi_1(U) \longrightarrow GL(V)$ where V is a finite dimensional $\bar{\mathbb{Q}}_\ell$-vector space. Continuity refers to the profinite topology on $\pi_1(U)$ and the ℓ-adic topology on V.

Note the similarity with the definition of Galois representations of number fields (see Section 5.13). Because of the original definition of a sheaf, one usually denotes ℓ-adic sheaves by curly letters $\mathcal{F}$, $\mathcal{G}$, etc. Notice that one can obviously speak of direct sums, tensor product, symmetric powers, etc., of lisse ℓ-adic sheaves by performing the corresponding operations on the representations. Also one can speak of irreducible sheaves, etc.

An important ℓ-adic sheaf, denoted $\bar{\mathbb{Q}}_\ell(1)$, is obtained by considering the natural action of $\pi_1(U)$ on ℓ-power roots of unity, which arises from the étale coverings where one simply extends the base field from $\mathbb{F}$ to its extension by roots of unity. This action is given by a certain character $\chi_\ell : \pi_1(U) \to \bar{\mathbb{Q}}_\ell^*$. Using this sheaf, one defines Tate twists: if $\mathcal{F}$ is a lisse ℓ-adic sheaf and $i \in \mathbb{Z}$, then one denotes $\mathcal{F}(i)$ ($\mathcal{F}$ twisted i times) the sheaf which corresponds to the action ρ' of $\pi_1(U)$ on the same vector space but with

$$\rho'(\gamma) = \chi_\ell^i(\gamma)\rho(\gamma);$$

in other words, $\mathcal{F}(1) = \mathcal{F} \otimes \bar{\mathbb{Q}}_\ell(1)$ for instance.

Exponential sums arise by looking at the action of the Frobenius elements at points of U. Let x be a closed point x of U, which can be seen as a Galois orbit of points in $U(\bar{\mathbb{F}})$. The fundamental group of the "point" x is the Galois group D_x of the residue field of U at x, isomorphic to $\mathbb{F}_n$ where n is the degree of x. By functoriality there is a map $D_x \to \pi_1(U)$. We have $D_x \simeq \mathrm{Gal}(\bar{\mathbb{F}}_n/\mathbb{F}_n)$ and the latter is generated (topologically) by the Frobenius morphism σ, so taking the image we get in $\pi_1(U)$ a well-defined conjugacy class, called the arithmetic Frobenius conjugacy class at x. In particular, for any ℓ-adic sheaf one can speak of the trace $\mathrm{Tr}\,\rho(\sigma_x)$ without ambiguity. However, it turns out that it is the inverse F of σ (the so-called geometric Frobenius) which appears naturally in the cohomological description of exponential sums. We denote by F_x the corresponding conjugacy class; it is called simply the Frobenius conjugacy class at x (omitting the adjective geometric).

THEOREM 11.34. *Let $U/\mathbb{F}$ be a smooth variety, let $f \neq 0$ and g be $\mathbb{F}$-rational functions on U, let ψ be an additive character and χ a multiplicative character of $\mathbb{F}$. Let $S_n = S_n(U, f, g, \chi, \psi)$ be the associated exponential sums over $U(\mathbb{F}_n)$ as in (11.43). Then there exists a lisse ℓ-adic sheaf $\mathcal{F}$ on U of degree 1 with the property that for all $n \geqslant 1$ we have*

$$S_n = \sum_{x \in U(\mathbb{F}_n)} Tr(F_x \mid \mathcal{F}) \tag{11.44}$$

where we denote $Tr(g \mid \mathcal{F}) = Tr(\rho(g) \mid V)$, $\mathcal{F}$ corresponding to the representation $\rho : \pi_1(U) \to GL(V)$.

To compare with the characters used to describe Gauss sums and Kloosterman sums, one should think of the latter as analogues of Dirichlet characters or Hecke characters, whereas the ℓ-adic sheaves given by this theorem are analogues of Galois characters. The correspondence between the two concepts is an instance of reciprocity or class-field theory.

We sketch the construction of $\mathcal{F}$ in the case where $\chi = 1$ and g is a non-zero rational function on $U = \mathbb{A}^1 - \{\text{poles of } g\}$, over $\mathbb{F}$, which makes it clear that this is very closely related to the argument in Section 11.7. Consider the curve

$$C : y^q - y = g(x) \tag{11.45}$$

and notice that there is a surjective map $\pi : (x, y) \mapsto x$ from C to U. For any $a \in \bar{\mathbb{F}}$, the equation $y^q - y - a = 0$ is separable, hence it has q distinct roots in $\bar{\mathbb{F}}$. In fact the additive group of $\mathbb{F}$ acts on the roots by translation: if y is a root and $z \in \mathbb{F}$, then $(y+z)^q - (y+z) = y^q - y = a$. Moreover, $\pi : C \to U$ is an étale covering (we've just seen it is everywhere unramified and surjective). In other words, π is an étale Galois covering with Galois group isomorphic to the additive group $\mathbb{F}$ (coverings given by such equations are called Artin-Shreier coverings).

The fundamental group $\pi_1(U)$ acts on C by automorphisms of the covering, which means as translations by elements of $\mathbb{F}$ as above. This defines a surjective map $\varphi : \pi_1(U) \to \mathbb{F}$ such that $\varphi(\gamma) = \gamma y - y$ for any $y \in C$. (This doesn't depend on the choice of y because the action of γ on C must be a morphism of curves.)

Consider the trivial ℓ-adic sheaf $\bar{\mathbb{Q}}_\ell$ on C, or equivalently the trivial representation of $\pi_1(C)$. By the above we can construct the induced representation ρ from

$\pi_1(C)$ to $\pi_1(U)$, which can be described as the space

$$V = \{f : \pi_1(U) \to \bar{\mathbb{Q}}_\ell \mid f(\tau\gamma) = f(\gamma) \text{ for any } \tau \in \pi_1(C)\}$$

(where $\tau \in \pi_1(C)$ is seen through the map $\pi_1(C) \to \pi_1(U)$ coming from π), on which $\pi_1(U)$ acts by translation on the right

$$\rho(\gamma)f(\tau) = f(\tau\gamma).$$

The elements $f \in V$ depend only on $\pi_1(C)\backslash\pi_1(U) \simeq \mathbb{F}$ (i.e. on the automorphisms of the covering $C \to U$), which implies that $V \simeq \bar{\mathbb{Q}}_\ell^q$ is an ℓ-adic sheaf on U of degree q. The representation space V can be decomposed over the additive characters ψ of $\mathbb{F}$,

$$V = \bigoplus_\psi \mathcal{L}_\psi$$

where $\mathcal{L}_\psi$ is the ψ-eigencomponent of V, namely

$$\mathcal{L}_\psi = \{f \in V \mid \rho(\gamma)f = \psi(\varphi(\gamma))f \text{ for all } \gamma \in \pi_1(U)\}.$$

It is easy to see that each $\mathcal{L}_\psi$ is an ℓ-adic sheaf on U, and because ρ is induced from the trivial representation, each $\mathcal{L}_\psi$ is of degree 1.

Then for every additive character ψ, the ℓ-adic sheaf on U corresponding to $\mathcal{L}_{\bar\psi}$ is the sheaf satisfying (11.44) for the exponential sums $S_n(U, g, \psi)$.

Indeed, if $x \in U(\mathbb{F}_n)$, and y satisfies $y^q - y = g(x)$, then the Frobenius of x acts on y by $y^{q^n} = y + \operatorname{Tr}_{\mathbb{F}_n/\mathbb{F}}(g(x))$ since

$$\begin{aligned} y^{q^n} - y &= y^{q^n} - y^{q^{n-1}} + y^{q^{n-1}} - \cdots + y^q - y \\ &= (y^q - y)^{q^{n-1}} + \cdots + y^q - y = \operatorname{Tr}(y^q - y) = \operatorname{Tr} g(x). \end{aligned}$$

Hence $\varphi(\sigma_x) = \operatorname{Tr} g(x)$ and by definition of $\mathcal{L}_\psi$ it follows that σ_x acts on $\mathcal{L}_\psi$ by multiplication by $\psi(\operatorname{Tr} g(x))$, hence $F_x = \sigma_x^{-1}$ acts by $\bar\psi(\operatorname{Tr} g(x))$, which gives (11.44).

In particular, note that taking the trace for $\mathbb{Q}_\ell$ on C we derive

$$|C(\mathbb{F}_n)| = \sum_\psi S_n(U, f, \psi),$$

as in (11.30).

Exercise 3. (1) Let S_n be the character sum (11.8) with $g = 0$ for some multiplicative character χ of $\mathbb{F}^\star$ and some non-zero rational function $f \in \mathbb{F}(x)$, on the variety $U = \mathbb{A}^1 - \{\text{zeros and poles of } f\}$. Describe as above the construction of the sheaf $\mathcal{L}$ satisfying (11.44) in this case. [**Hint:** Use the cover $y^d = f(x)$, where d is the order of the multiplicative character χ.]

(2) Let S_n be as in (11.8), $U \subset \mathbb{A}^1$ the complement of the zeros and poles of f and the poles of g. If $\mathcal{L}_\psi$ is the sheaf satisfying (11.44) for $f = 1$ and $\mathcal{L}_\chi$ is the sheaf satisfying (11.44) for $g = 0$, show that $\mathcal{L} = \mathcal{L}_\psi \otimes \mathcal{L}_\chi$ satisfies (11.44) for S_n.

Examples. (1) Even the case $\rho = 1$ is interesting when dealing with a general variety U. This "trivial" ℓ-adic sheaf is denoted $\bar{\mathbb{Q}}_\ell$, and one has $S_n = |U(\mathbb{F}_n)|$.

(2) For the sheaf $\bar{\mathbb{Q}}_\ell(1)$, notice that σ_x acts by $\xi \mapsto \xi^q$ for any root of unity if $Nx = q$. Therefore F_x acts by $\xi \mapsto \xi^{1/q}$ and in particular the only eigenvalue of F_x is q^{-1}.

Now in addition to $U/\mathbb{F}$ we consider its "extension of scalars" $\bar{U}/\bar{\mathbb{F}}$ over the algebraic closure of $\mathbb{F}$. There is a corresponding geometric fundamental group $\pi_1(\bar{U})$, which sits in an exact sequence

$$1 \to \pi_1(\bar{U}) \to \pi_1(U) \to \mathrm{Gal}(\bar{\mathbb{F}}/\mathbb{F}) \to 1. \tag{11.46}$$

To every ℓ-adic sheaf $\mathcal{F}$ on U are associated the ℓ-adic cohomology groups with compact support of $\bar{U}$ with coefficients in $\mathcal{F}$. Those are finite-dimensional $\bar{\mathbb{Q}}_\ell$-vector spaces, denoted, $H^i_c(\bar{U}, \mathcal{F})$ for $i \geqslant 0$. The key point is that the Galois group of $\mathbb{F}$ acts naturally on $H^i_c(\bar{U}, \mathcal{F})$, and in particular, so do the Frobenius σ and its inverse F, the geometric Frobenius. The key to the cohomological interpretation of exponential sums is the

Grothendieck-Lefschetz Trace Formula. *Let $U/\mathbb{F}$ be a smooth variety of dimension $d \geqslant 0$, $\mathcal{F}$ an ℓ-adic sheaf on U. We have $H^i_c(\bar{U}, \mathcal{F}) = 0$ if $i > 2d$ and for any $n \geqslant 1$*

$$\begin{aligned}\sum_{x \in U(\mathbb{F}_n)} Tr(F_x \mid \mathcal{F}) = {} & Tr(F^n \mid H^0_c(\bar{U}, \mathcal{F})) - Tr(F^n \mid H^1_c(\bar{U}, \mathcal{F})) + \cdots \\ & - Tr(F^n \mid H^{2d-1}_c(\bar{U}, \mathcal{F})) + Tr(F^n \mid H^{2d}_c(\bar{U}, \mathcal{F})).\end{aligned} \tag{11.47}$$

Therefore to evaluate the exponential sums (11.43) using the associated sheaf, we need to know the traces, or equivalently the eigenvalues, of F (equivalently, of $\sigma = F^{-1}$) acting on H^i_c for $0 \leqslant i \leqslant 2d$. It turns out that in most cases H^0_c and H^{2d}_c are easy to compute:

Proposition 11.35. *Let $\mathcal{F}$ be a lisse ℓ-adic sheaf on a smooth variety $U/\mathbb{F}$, corresponding to the representation ρ of $\pi_1(U)$ on the $\bar{\mathbb{Q}}_\ell$-vector space V. We have*

$$H^0_c(\bar{U}, \mathcal{F}) \simeq \begin{cases} V^{\pi_1(\bar{U})} & \text{if } U \text{ is projective,} \\ 0 & \text{if } U \text{ is not projective,} \end{cases} \tag{11.48}$$

and

$$H^{2d}_c(\bar{U}, \mathcal{F}) \simeq V_{\pi_1(\bar{U})}(-2d) \tag{11.49}$$

where V^G denotes the space of vectors invariant under the action of a group G on an abelian group, and V_G denotes the space of co-invariants, the largest quotient of V on which G acts trivially. In both cases, the isomorphisms are canonical isomorphisms of vector spaces with an action of the Galois group of $\mathbb{F}$.

Since V is a representation of $\pi_1(U)$, the exact sequence (11.46) shows that $V^{\pi_1(\bar{U})}$ and $V_{\pi_1(\bar{U})}$ are acted on by $\mathrm{Gal}(\bar{\mathbb{F}}/\mathbb{F})$, "through" the given representation ρ.

This proposition shows that for a curve $U/\mathbb{F}$, the only "difficult" cohomology group is $H^1_c(\bar{U}, \mathcal{F})$.

Example. Let $U = E/\mathbb{F}_p$ be an elliptic curve, $\mathcal{F} = \bar{\mathbb{Q}}_\ell$ the trivial sheaf. By the proposition one has

(1) $H^0_c(\bar{E}, \bar{\mathbb{Q}}_\ell) = \bar{\mathbb{Q}}_\ell$, with trivial action of F (since $\bar{\mathbb{Q}}_\ell$ is the trivial sheaf).

(2) $H^2_c(\bar{E}, \bar{\mathbb{Q}}_\ell) = \bar{\mathbb{Q}}_\ell(-1)$, so by definition of the twist, F acts by multiplication by p (on roots of unity, i.e. on $\bar{\mathbb{Q}}_\ell(1)$, σ acts by $\xi \mapsto \xi^p$, hence F by multiplication by p^{-1}).

The Lefschetz trace formula (11.47) gives

$$|E(\mathbb{F}_n)| = p^n + 1 - \operatorname{Tr}\left(F^n \mid H_c^1(\bar{E}, \bar{\mathbb{Q}}_\ell)\right)$$

(compare Theorem 11.31).

More generally, one derives the rationality of the zeta function directly from the Trace Formula.

COROLLARY 11.36. *Let U, S_n and $\mathcal{F}$ be as in Theorem 11.34. For $0 \leqslant i \leqslant 2d$, let $b_i = \dim H_c^i(\bar{U}, \mathcal{F})$ and*

$$P_i(T) = \det(1 - FT \mid H_c^i(\bar{U}, \mathcal{F})) = \prod_{j=1}^{b_i} (1 - \alpha_{i,j}T).$$

We have

$$Z(\mathcal{F}) = \exp\Big(\sum_{n \geqslant 1} \frac{S_n}{n} T^n\Big) = \frac{P_1(T) \cdots P_{2d-1}(T)}{P_0(T) \cdots P_{2d}(T)} = \prod_{i=0}^{2d} \prod_j (1 - \alpha_{i,j}T)^{(-1)^{i+1}},$$

and for $n \geqslant 1$,

$$S_n = \sum_{0 \leqslant i \leqslant 2d} (-1)^i \alpha_{i,j}^n. \tag{11.50}$$

Theorems 11.4, 11.8 and the result of Exercise 1 are all special cases of this corollary, together with suitable computations of cohomology groups. The numbers b_i are called the ℓ-adic Betti numbers for $\mathcal{F}$.

Of much greater importance, however, is Deligne's vast generalization of the Riemann Hypothesis [**De2**]. One starts with the following "local" definition:

DEFINITION. Let $w \in \mathbb{Z}$ be an integer. A lisse ℓ-adic sheaf $\mathcal{F}$ on $U/\mathbb{F}$ is said to be pure of weight w if for any closed point x of U, all eigenvalues of F_x acting on the $\bar{\mathbb{Q}}_\ell$ vector space V associated to $\mathcal{F}$ are algebraic numbers all conjugates of which have the same absolute value equal to $q^{w/2}$ where $q = Nx$ is the cardinality of the residue field.

For instance, the trivial sheaf $\bar{\mathbb{Q}}_\ell$ is pure of weight 0 (all eigenvalues 1). For any $i \in \mathbb{Z}$, $\bar{\mathbb{Q}}_\ell(i)$ is pure of weight $-2i$, and if $\mathcal{F}$ is pure of weight w, then $\mathcal{F}(i)$ is pure of weight $w - 2i$. For any exponential sum (11.43), the associated sheaf $\mathcal{F}$ is pure of weight 0 because the only eigenvalue at x is the root of unity $\chi(Nf(x))\psi(\operatorname{Tr} g(x))$.

THEOREM 11.37 (DELIGNE). *Let $U/\mathbb{F}$ be a smooth variety and $\mathcal{F}$ a lisse ℓ-adic sheaf on U, pure of weight w. Let $i \geqslant 0$ and let ξ be any eigenvalue of the geometric Frobenius F acting on $H_c^i(\bar{U}, \mathcal{F})$. Then ξ is an algebraic integer, and if $\alpha \in \mathbb{C}$ is a conjugate of ξ, we have*

$$|\alpha| \leqslant q^{(w+i)/2}. \tag{11.51}$$

The conclusion is also phrased as saying that $H_c^i(\bar{U}, \mathcal{F})$ is mixed of weights $\leqslant i + w$. If there is equality in (11.51), then $H_c^i(\bar{U}, \mathcal{F})$ is said to be pure (of weight $w + i$). In certain cases, one can apply duality theorems (for instance Poincaré duality) to deduce further that (11.51) is an equality.

REMARK. Although Deligne's proof is a monumental achievement of very deep algebraic geometry, it is an interesting fact that a crucial use is made of a generalization of the method of Hadamard and de la Vallée Poussin for proving non-vanishing of L-functions on the line $\mathrm{Re}\,(s) = 1$ (see Section 5.4). Similarly, in Deligne's first proof [**De1**], the ideas of the classical Rankin-Selberg method for modular forms are essential (specifically, Deligne acknowledges the influence of [**Ra3**]).

EXAMPLE. Let $C/\mathbb{F}$ be a smooth connected projective curve (for instance, an elliptic curve). By Proposition 11.35 as in the previous example, we have easily:

(1) $H_c^0(\bar{C}, \bar{\mathbb{Q}}_\ell) = \bar{\mathbb{Q}}_\ell$, with F acting trivially.

(2) $H_c^2(\bar{C}, \bar{\mathbb{Q}}_\ell) = \bar{\mathbb{Q}}_\ell(-1)$, with F acting by multiplication by p.

It is more difficult to show that

(3) $H_c^1(\bar{C}, \bar{\mathbb{Q}}_\ell) \simeq \bar{\mathbb{Q}}_\ell^{2g}$, as $\bar{\mathbb{Q}}_\ell$ vector spaces (not as Galois-modules!), where $g \geqslant 0$ is the genus of $\bar{C}$ (for an elliptic curve $g = 1$).

Moreover, there is a Galois-invariant perfect pairing

$$H_c^1(\bar{C}, \bar{\mathbb{Q}}_\ell) \times H_c^1(\bar{C}, \bar{\mathbb{Q}}_\ell) \longrightarrow \bar{\mathbb{Q}}_\ell(-1).$$

It follows that if α is one of (the complex conjugates of) the eigenvalues of F on $H_c^1(\bar{C}, \bar{\mathbb{Q}}_\ell)$, then p/α is one also. Hence from Theorem 11.37, since $\bar{\mathbb{Q}}_\ell$ is pure of weight 0, one deduces that $|\alpha| = \sqrt{p}$. Thus

$$|C(\mathbb{F}_n)| = p^n + 1 - \sum_{i=1}^{2g} \alpha_i^n$$

where the $\alpha_i \in \bar{\mathbb{Q}}$ are the eigenvalues of F on H_c^1. Estimating trivially now, we get

$$\Big||C(\mathbb{F}_n)| - (p^n + 1)\Big| \leqslant 2gp^{n/2}$$

recovering the Riemann Hypothesis, and in particular, Theorem 11.25 for the case $g = 1$.

In the case of exponential sums (11.43), the sheaf $\mathcal{F}$ is pure of weight 0, hence denoting

$$d(\mathcal{F}) = \max\{i \mid H_c^i(\bar{U}, \mathcal{F}) \neq 0\},$$

we derive directly from (11.50) and (11.51) the bound

$$|S_n| \leqslant \sum_{0 \leqslant i \leqslant d(\mathcal{F})} b_i q^{ni/2}, \tag{11.52}$$

for $n \geqslant 1$ and, in particular,

$$|S_n| \leqslant q^{nd(\mathcal{F})/2}\Big(\sum_i b_i\Big). \tag{11.53}$$

As in the case of Kloosterman sums, the exponent $d(\mathcal{F})/2$ is best possible in this inequality. The bound $d(\mathcal{F}) \leqslant 2d$ gives a trivial estimate (because U is smooth of dimension d, it has about q^{nd} points, as proved by the Riemann Hypothesis for the trivial sheaf $\bar{\mathbb{Q}}_\ell$). Any improvement of this trivial bound is equivalent with $H_c^{2d}(\bar{U}, \mathcal{F}) = 0$, and the square root cancellation often expected from heuristic reasonings is equivalent with $H_c^i(\bar{U}, \mathcal{F}) = 0$ for $i > d$. Although not always true, this turns out to hold "generically", as the analytic intuition suggests (see for instance Theorem 11.43 below).

For the exponential sums (11.43) we have $d(\mathcal{F}) < 2d$, unless $\mathcal{F}$ is the trivial sheaf $\bar{\mathbb{Q}}_\ell$, so there is always a non-trivial bound. This follows from (11.49), since $\mathcal{F}$ is of degree 1 so the space of co-invariants is either the whole space (meaning the representation is trivial) or 0. However, this small gain is usually insufficient in applications.

Another surprising consequence of Deligne's result and the discreteness of integers is the following "self-improving" statement:

COROLLARY 11.38. *Let S_n be an exponential sum as in (11.43) and $\mathcal{F}$ the associated sheaf. Suppose $w \geqslant 0$ is an integer such that*

$$|S_n| \ll q^{w/2+\delta}$$

for some $\delta \in [0, \frac{1}{2}[$ and $n \geqslant 1$. Then we have $d(\mathcal{F}) \leqslant w$, hence $|S_n| \ll q^{w/2}$.

A second issue in applying the estimates (11.52) or (11.53) in the context of applications to analytic number theory is that we usually have $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$, with the prime number p varying. In this case, whereas the variety U can be defined over $\mathbb{Q}$ (or $\mathbb{Z}$) so that the sum is, for all p, over the $\mathbb{F}_p$-points of the reduction U_p of U modulo p, the sheaves $\mathcal{F}_p$ genuinely depend on p (see the equation (11.45)), i.e. there is no theory of sheaves over $U/\mathbb{Z}$ giving each $\mathcal{F}_p$ by "reduction modulo p". (Katz has asked a number of times for such a theory of "exponential sums over $\mathbb{Z}$"; see e.g. [**K2**], but it remains elusive.) Thus, the Betti numbers

$$b_i(p) = \dim H_c^i(\bar{U}_p, \mathcal{F}_p)$$

of the cohomology groups can depend on p, and the applicability of the results above would be ruined, even with the Riemann Hypothesis, if these dimensions were not bounded in a reasonable way in terms of p.

This is in fact the case. The first general result in this direction is due to Bombieri [**Bo4**] for additive character sums (11.43) where $f = 1$, and was generalized by Adolphson and Sperber [**AS1**], [**AS2**] for general sums (their methods are p-adic, based on Dwork's original ideas). In general, those results bound the Euler characteristic

$$\chi_c(\mathcal{F}) = \sum_{i=0}^{2d} (-1)^i \dim H_c^i(\bar{U}, \mathcal{F}) = \sum_{0 \leqslant i \leqslant 2d} (-1)^i b_i,$$

of a sheaf $\mathcal{F}$ on $U/\mathbb{F}$, but further arguments of Katz [**K5**] show how to deduce bounds for

$$\sigma_c(\mathcal{F}) = \sum_{i=0}^{2d} \dim H_c^i(\bar{U}, \mathcal{F}) = \sum_{0 \leqslant i \leqslant 2d} b_i,$$

(hence for $b_i \leqslant \sigma_c(\mathcal{F})$) from those for $\chi_c(\mathcal{F})$.

THEOREM 11.39. *Let $U/\mathbb{Q}$ be a smooth variety over $\mathbb{Q}$, f and g functions on U with f invertible. Let ℓ be a prime number and for all $p \neq \ell$ such that the reduction U_p of U modulo p is smooth, let χ and ψ be any multiplicative and additive characters of $\mathbb{F}_p$. Let $\mathcal{F}_p$ be an ℓ-adic sheaf on U_p such that*

$$\sum_{x \in U(\mathbb{F}_{p^n})} \chi(Nf(x))\psi(Trg(x)) = \sum_{x \in U(\mathbb{F}_{p^n})} Tr(F_x \mid \mathcal{F}_p)$$

for $n \geqslant 1$. We have $\sigma_c(\mathcal{F}_p) \leqslant C$ where C is a constant depending only on U, f and g.

A simple explicit bound is given in [**AS3**] if $f(x) = 1$, so only the additive characters occur, and g is a Laurent polynomial on $U = (\bar{\mathbb{Q}} - \{0\})^d$. The sums over $\mathbb{Z}/p\mathbb{Z}$ in question are therefore sums in d variables of the type

$$S_{f,p} = \sum_{x_1,\dots,x_d \in (\mathbb{Z}/p\mathbb{Z})^\star} \psi(f(x_1,\dots,x_d)) \tag{11.54}$$

where $f \in \mathbb{Q}[x_1, x_1^{-1}, \dots, x_d, x_d^{-1}]$ is a non-zero Laurent polynomial. Writing

$$f = \sum_{j \in J} a_j x^j$$

for some (finite) set $J \subset \mathbb{Z}^d$, the Newton polyhedron $W(f)$ of f is defined to be the convex hull in $\mathbb{R}^d$ of $J \cup \{0\}$.

PROPOSITION 11.40. *With the above assumptions, denoting by $\mathcal{F}_{f,p}$ the associated sheaf for the sums $S_{f,p}$, we have*

$$\begin{aligned} |\chi_c(\mathcal{F}_{f,p})| &\leqslant d!\mathrm{Vol}\,(W(f)), \\ \sigma_c(\mathcal{F}_{f,p}) &\leqslant 10^d d!\mathrm{Vol}\,(W(f)) \end{aligned}$$

for any p not dividing the denominator of any coefficient of f, where $\mathrm{Vol}\,(W(f))$ is the volume of the Newton polyhedron in the subspace spanned by $W(f)$ in $\mathbb{R}^d$, with respect to Lebesgue measure.

Note that by using exclusion-inclusion and detecting polynomials equations by means of multiplicative characters, one can use combinations of sums of the type (11.54) to describe much more general ones. Also, in many cases, one can show that all the odd (or even) cohomology groups vanish, in which case $|\chi_c(\mathcal{F})| = \sigma_c(\mathcal{F})$. See also Theorems 11 and 12 of [**K5**] for explicit estimates in quite general cases.

We now give examples of computations using these fundamental results. For exponential sums arising in analytic number theory, one often needs nothing more, if one uses skillfully some other simple tricks such as averaging over extra parameters to analyze the weight of the roots.

EXAMPLE 1. The Kloosterman sums $S(a,b;p)$ for $ab \neq 0$ can be treated using Proposition 11.40 with $d = 1$ and $f(x) = ax + bx^{-1}$. Then $W(f)$ is the interval $[-1,1]$. By Proposition 11.35, we have $H_c^0 = H_c^2 = 0$ in this case since $U = \mathbb{P}^1 - \{0,\infty\}$ is not projective, so $\sigma_c = -\chi_c$. By Theorem 11.37, H_c^1 is mixed of weight $\leqslant 1$. Hence we recover the Weil bound:

$$|S(a,b;p)| \leqslant \sigma_c p^{1/2} \leqslant 2p^{1/2}.$$

(Of course, in fact we have $b_1 = 2$ and the last inequality is an equality).

EXAMPLE 2. The previous example generalizes to the multiple Kloosterman sums defined by

$$K_r(a,q) = \sum_{x_1 \cdots x_r = a} e\Big(\frac{\mathrm{Tr}\,(x_1 + \cdots + x_r)}{p}\Big) \tag{11.55}$$

for $r \geqslant 2$ and $a \neq 0$, so $K_2(a,p) = S(a,1;p)$ (see [**Bo4**], [**De1**]). Without appealing to the L-function, one can nevertheless get some information by averaging over a. We get

$$\sum_{a\neq 0} |K_r(a,q)|^2 = q^r - q^{r-1} - \cdots - q - 1. \tag{11.56}$$

Hence $|K_r(a,q)| \leqslant q^{r/2}$. To improve this elementary bound we appeal to the following Lemma

LEMMA 11.41. *Given a finite set of distinct angles θ_i modulo 2π and complex numbers α_i we have*

$$\sum_{n\leqslant N} \Big|\sum_i \alpha_i e(n\theta_i)\Big|^2 = N\|\alpha\|^2 + O(1)$$

where the implied constant does not depend on N. Hence

$$\limsup_{n\to+\infty} \Big|\sum_i \alpha_i e(n\theta_i)\Big| \geqslant \|\alpha\|. \tag{11.57}$$

PROOF. We have

$$\sum_{n\leqslant N} \Big|\sum_i \alpha_i e(n\theta_i)\Big|^2 = N\sum_i |\alpha_i|^2 + \sum_{i\neq j}\sum \alpha_i\alpha_j \sum_{n\leqslant N} e(n(\theta_i-\theta_j)).$$

The inner sum is bounded by a constant independent of N, so the first result follows and (11.57) is an obvious consequence. □

From (11.56) and (11.57) it follows that among the $K_r(a,q)$, there is at most one root of weight r, say for $K_r(a_0,q)$, and all other roots are of weight $\leqslant r-1$.

Notice that $K_r(a_0,q) \in \mathbb{Q}(\mu_p)$, the cyclotomic field of p-th roots of unity. Using the Galois action on $\mathbb{Q}(\mu_p)$, the conjugates of $K_r(a_0,q)$ are $K_r(a_0v^r,q)$ for $v \in \mathbb{F}_p^\star$. By the Riemann Hypothesis, this means that the conjugate of the root ξ of weight r is still a root of weight r for $K_r(a_0v^r,q)$. Hence $v^r = 1$ for all $v \in \mathbb{F}_p^\star$, which is only possible if $p-1 \mid r$. In particular all roots are of weight $\leqslant r-1$ if $p > r+1$. One therefore gets by Proposition 11.40

$$K_r(a,q) \ll q^{(r-1)/2} \tag{11.58}$$

where the implied constant depends only on r.

In the case of Kloosterman sum the Newton polyhedron is the simplex with vertices $(1,0,\dots,0)$, $\dots$, $(0,\dots,0,1)$, $(-1,\dots,-1)$ whose volume is $1/r!$. Moreover, it is known that the zeta function is a polynomial so $\chi_c = -\sigma_c$ and we get the precise estimate

$$|K_r(a,q)| \leqslant rq^{(r-1)/2}.$$

This was first proved by Deligne [**De3**], without any assumption on p and r.

EXAMPLE 3. Here is another higher-dimensional example. In [**CI1**], the following exponential sum over finite fields appears:

$$W(\chi,\psi;p)=\sum_{x,y \bmod p}\chi(xy(x+1)(y+1))\psi(xy-1) \tag{11.59}$$

where p is prime, χ is a non-trivial quadratic character modulo p and ψ is any multiplicative character modulo p.

THEOREM 11.42 (CONREY-IWANIEC). *There exists an absolute positive constant C such that*

$$|W(\chi,\psi;p)|\leqslant Cp \tag{11.60}$$

for all p and all ψ as above.

The first step of the proof is to apply Theorem 11.34 to say there exists an ℓ-adic sheaf $\mathcal{F}$ on the algebraic surface

$$U=\{(x,y)\mid xy(x+1)(y+1)\neq 0\},$$

pure of weight 0 for which we have, for $q=p^n$, $n\geqslant 1$, the formula

$$\begin{aligned}W(\chi,\psi;q)&=\sum_{x,y\in U(\mathbb{F}_q)}\chi(N(xy(x+1)(y+1)))\psi(N(xy-1))\\&=\sum_{x,y\in U(\mathbb{F}_q)}\mathrm{Tr}\,(F_{(x,y)}\mid\mathcal{F}).\end{aligned}$$

The Lefschetz trace formula (11.47) takes the form

$$W(\chi,\psi;q)=\sum_{i=0}^{4}(-1)^i\mathrm{Tr}\,(F^n\mid H^i_c(\bar{U},\mathcal{F})).$$

By Theorem 11.37, each $H^i_c(\bar{U},\mathcal{F})$ is mixed of weights $\leqslant i$. Let (α_ν,i_ν,w_ν) be the family of eigenvalues of F acting on the whole cohomology (with multiplicities), together with their index and weight. We have $|\alpha_\nu|=p^{w_\nu/2}$ and

$$W(\chi,\psi;q)=\sum_\nu(-1)^{i_\nu}\alpha_\nu^m.$$

By Theorem 11.39 in this case, the total number of roots α_ν is bounded by a constant independent of p. Thus, we gain on the trivial bound $W\ll p^2$ if $w_\nu\leqslant 3$ (instead of 4), and the statement of the theorem is that $w_\nu\leqslant 2$.

The second step is to show that there is at most one root of weight $\geqslant 3$ (actually, it must be $=3$) and, if it exists, then $\psi=\chi$ is the non-trivial quadratic character. This will be derived from the following average formula:

$$A=\frac{1}{q-1}\sum_\psi|W(\chi,\psi;q)|^2=q^2-2q-2. \tag{11.61}$$

To prove this formula, open the square and sum over ψ first getting by the orthogonality of characters

$$\begin{aligned} A &= \sum_{u_1 v_1 = u_2 v_2} \sum \chi(u_1 v_1 (u_1+1)(v_1+1)) \bar{\chi}(u_2 v_2 (u_2+1)(v_2+1)) \\ &= \sum_{u_1, v_1, u_2} \sum \chi((u_1+1)(v_1+1)) \bar{\chi}((u_2+1)(u_1 v_1 + u_2)) \chi(u_2) \end{aligned}$$

(where we shorten the notation from $\chi \circ N$ to χ). Next the summation over u_1 is performed giving

$$B(v_1, u_2) = \sum_{u_1 \neq 0} \chi(u_1+1) \bar{\chi}(u_1 v_1 + u_2) = \begin{cases} \bar{\chi}(v_1)(q-2) & \text{if } u_2 = v_1, \\ -\bar{\chi}(v_1) - \bar{\chi}(u_2) & \text{if } u_2 \neq v_1. \end{cases}$$

Then the sum over u_2 is performed giving

$$\sum_{u_2 \neq 0} \bar{\chi}(u_2+1) \chi(u_2) B(v_1, u_2) = q \bar{\chi}(v_1+1) + \bar{\chi}(v_1) + 1,$$

and finally the sum over v_1 gives

$$A = \sum_{v_1 \neq 0} \chi(v_1+1) \Big(q \bar{\chi}(v_1+1) + \bar{\chi}(v_1) + 1 \Big) = q(q-2) - 2.$$

From (11.61), using Lemma 11.41, it is clear that for all ψ and ν we have $w_\nu \leqslant 3$ and moreover, $w_\nu \leqslant 2$ except for at most one root, for one character ψ. If this case occurs for ψ, it happens for $\bar{\psi}$ too, so the only possibility is ψ being a real character. Since

$$W(\chi, 1; q) = \Big(\sum_u \chi(u(u+1)) \Big)^2 = 1,$$

we must have $\psi = \chi$.

The last step is to treat the case $\psi = \chi$ separately. Precisely, one can show that

$$|W(\chi, \chi; q)| \leqslant 4q$$

for any p and $q = p^n$. This is done in a purely elementary manner without appealing to the Riemann Hypothesis, and we refer to [**CI1**] for the details. In fact, W. Duke showed that $W(\chi, \chi; p) = 2\mathrm{Re}\,(J^2(\chi, \xi))$, where $J(\chi, \xi)$ is the Jacobi sum and ξ is a quartic character modulo $p \equiv 1 \pmod 4$. Also $W(\chi, \chi; p)$ is the p-th Fourier coefficient of the modular form $\eta(4z)^6$ of weight 3 and level 12.

When U is not a curve, numerous geometric subtleties can be involved in dealing with the non-trivial cohomology groups H_c^i with $i \neq 0,\ 2d$. Here are two general bounds, among many: the first one is due to Deligne [**De1**], and the second is the recent version in [**FK**] of a general "stratification" theorem of Katz and Laumon.

THEOREM 11.43. *(1) Let $f \in \mathbb{Z}[X_1, \dots, X_m]$ be a non-zero polynomial of degree d such that the hypersurface H_f in $\mathbb{P}^{m-1}$ defined by the equation*

$$H_f \ : \ f_d(x_1, \dots, x_m) = 0,$$

where f_d is the homogeneous component of degree d of f, is non-singular. For any $p \nmid d$ such that the reduction of H_f modulo p is smooth, any non-trivial additive character ψ modulo p and any $n \geqslant 1$ we have

$$\Big|\sum_{x_1,\dots,x_m\in\mathbb{F}_{p^n}}\cdots\sum \psi(Tr(f(x_1,\dots,x_m)))\Big| \leqslant (d-1)^m q^{nm/2}. \tag{11.62}$$

(2) Let $d \geqslant 1$, $n \geqslant 1$ be integers, V a locally closed subscheme of $\mathbb{A}^n_{\mathbb{Z}}$ of dimension $\dim V(\mathbb{C}) \leqslant d$ and $f \in \mathbb{Z}[X_1,\dots,X_n]$ a polynomial.

Then there exists $C = C(n,d,V,f)$ and closed subschemes $X_j \subset \mathbb{A}^n_{\mathbb{Z}}$ of relative dimension $\leqslant n-j$ such that

$$X_n \subset \dots \subset X_2 \subset X_1 \subset \mathbb{A}^n_{\mathbb{Z}}$$

and for any rational function g non-zero on V, any $h \in (\mathbb{Z}/p\mathbb{Z})^n - X_j(\mathbb{Z}/p\mathbb{Z})$, any prime p, any non-trivial additive character ψ and multiplicative character χ modulo p, we have

$$\sum_{x\in V(\mathbb{Z}/p\mathbb{Z})} \chi(g(x))\psi(f(x)+h_1x_1+\cdots h_nx_n) \leqslant Cp^{\frac{d}{2}+\frac{j-1}{2}}.$$

Note that in (1), subject to a geometric condition on H_f, we obtain square root cancellation in the exponential sum. In (2) the assumptions are much less stringent, but the conclusion is weaker: we have a family of exponential sums parameterized by $h \in \mathbb{A}^n$, and roughly speaking we have square root cancellation for "generic" sums (for h outside an exceptional subvariety X_1 of codimension $\geqslant 1$ in $\mathbb{A}^n$), while worse and worse bounds can occur only on smaller and smaller subvarieties. We will give an application of (2) in Chapter 21.

The ℓ-adic theory and formalism are much more developed than what we have surveyed here. It can also deal with sums over singular varieties, but the necessary algebraic notions become rather formidable, and we concede being unable to discuss the perverse sheaves that arise in more advanced situations.

We wish to emphasize, however, that this theory is also particularly well-suited to the study of *families* of exponential sums. The parameters defining those, say S_x, are most naturally themselves points on some algebraic variety $X/\mathbb{F}$. In favorable circumstances there exists a lisse ℓ-adic sheaf $\mathcal{F}$ on X (corresponding to an action of $\pi_1(X)$ on V) such that

$$S_{x,n} = \mathrm{Tr}\,(F_x^n \mid V) \tag{11.63}$$

for every value of the parameter $x \in X$ and any $n \geqslant 1$. For example a non-zero polynomial f of degree $\leqslant d$ over $\mathbb{F}$ can be described as an $\mathbb{F}$-rational point of the affine parameter space $X = \mathbb{A}^{d+1} - \{0\}$ by

$$f = \sum_{i=0}^{d} a_i X^i \mapsto (a_0,\dots,a_d).$$

There is an ℓ-adic sheaf $\mathcal{F}$ on X such that

$$\sum_{x\in\mathbb{F}_n} \psi(f(x)) = \mathrm{Tr}\,(F_x^n \mid V).$$

Purity of a sheaf satisfying (11.63) depends on a first application of the Riemann Hypothesis. If it holds, the application of the ℓ-adic theory typically results in equidistribution statements (following from [**De2**]) for the arguments of the exponential sums. This equidistribution is in some space of conjugacy classes of the "monodromy group" of the situation. We refer for instance to [**KS**] for a very lucid introduction to these profound aspects.

11.12. Comments.

In this closing section we give some impressions about how the exponential and character sums over finite fields interact with analytic number theory. There are more subtle issues between the two subjects than just applications of results concerning the first one for solving problems of the latter. We could be quite specific by covering completely a few representative examples, but it would be long and not transparent enough. Rather we decided to discuss principles, ideas and tricks in general terms and guide the reader to particular publications.

First of all some exponential sums appear when one uses Fourier analysis to get a hold on the sequence under investigation. There are no finite fields in the background, so the resulting sums are not immediately related to objects of algebraic geometry. However, one can complete these sums (by another use of Fourier analysis) and then factor them into sums of prime power moduli. Usually one can evaluate these local sums explicitly, or give strong estimates by elementary or ad hoc methods, except when the modulus is prime. But in the prime case one may naturally consider the sum as being over a finite field. This scheme allows us to appeal to the powerful results from algebraic geometry. However, the drawback of finishing by estimates for every complete sum individually is that one cannot exploit a possible cancellation from extra variables offered by the varying moduli (finite fields of different characteristics do not interact in algebraic geometry). Sometimes a kind of reciprocity formula can help turn the modulus into a variable (see for example [**I11**] or [**M3**]). Another scenario is that the sums over modulus appear in the spectral resolution of a differential operator, in which case the spectral theory produces estimates far stronger than those derived by algebraic geometry. For example, this is the case of sums of Kloosterman sums; see Chapter 16. One can also think this way about the real character sums with cubic polynomials; they are coefficients of a cusp form associated with an elliptic curves, so the modularity gives extra cancellation in summation over the modulus.

As a rule the exponential/character sum of a given modulus which comes out of analytic number theory is incomplete. This itself is not a problem because various completing techniques are available as mentioned above. Completing is a natural step to take, but is it useful or wasteful? At this point one should realize that a bound for a complete sum holds essentially the same for the original incomplete sum. This means that the result is relatively weaker for a shorter sum. Still it is non-trivial when the length of the original sum is larger than the square root of the modulus. Very short sums cannot be treated this way. We do not have an absolute recommendation when to complete or not a given incomplete sum. Our experience suggests executing the Fourier method as long as the resulting summation over the frequencies is shorter than the range of the original sum: at least one can feel that one is progressing. But sometimes it is worth acting otherwise, accepting a step backwards in this respect while opening a position for a stronger second move.

For example, imagine that the amplitude in the exponential sum is not a rational function, but nevertheless becomes one after an application of the stationary phase method to the relevant Fourier transform. In this case the losses from the range of summation can be recovered with extra savings by applications of algebraic arguments (see Section 8 of [**CI1**], where this game is played in several variables simultaneously).

Whatever the arguments which lead to complete sums may be, in the final step one cannot beat the square root saving factor. Therefore to receive a non-trivial result one must first produce somehow a sum with a number of terms larger than the square root of the modulus. There are several ways to get started, depending on the shape of the exponential/character sum. First, one can try to apply a Weyl shift with the effect of squaring the number of terms. Similarly one may just square the whole sum, or raise it to a higher power to produce even more points. Note that shifting the variable and squaring the sum are not the same things; the first requires some additive features of the variable while the latter nothing at all. These operations seem to be quite superficial at first glance (we are taking essentially replicas of the original sum), yet with ingenuity one can rearrange the points so the summation goes in a skewed direction, the consecutive terms repel violently and randomly producing a considerable cancellation. This is easy to say, much harder to execute. In fact one needs many other devices, such as gluing several variables with small multiplicity in order to arrive at a single variable over a range larger than the square root of the modulus. One also must smooth out this composite variable before applying algebraic arguments. Usually an application of Cauchy's inequality and enlarging the outer summation (due to positivity) does the job. A powerful example how this works is given by Burgess [**Bur1**]. In this paper a short character sum is estimated by an appeal to the Riemann Hypothesis for algebraic curves. An interesting point is that after all the tricks one comes to a complete character sum for a curve of a large genus, although the original sum is over a line segment.

Different arguments for building one extra large variable are applied in [**FI4**], consequently the final complete sum comes in three variables, or equivalently in four variables over a hypersurface. Here the Deligne theory applies (see the Appendix by Birch and Bombieri), although the related variety is singular. One should not be surprised and afraid of that singularity, because, after all, the process of creating more points of summation at the start is quite superficial. In this game one must be experienced when mixing the points to be sure that it is quite random. Another interesting case of creating and estimating exponential sums in three variables is given by N. Pitt [**Pi**].

Applications of the Riemann Hypothesis for curves over finite fields are by now customary. Much less successful are the use of genuinely higher-dimensional varieties. There are reasons for the difficulties involved. First of all when more variables appear, stronger restrictions are imposed on them which are harder to resolve (a kind of uncertainty principle). Just imagine having an abundance of points to work with, but which are not free because of some side conditions. For example how would you cope with a requirement that the determinants of a family of elliptic curves match the conductors?

Of course, there are also direct applications of the Riemann Hypothesis for varieties to traditional problems of solvability of diophantine equations by means of the circle method (see examples in Chapter 20). If the number of variables is sufficiently large, one needs nothing to manipulate, except for completing the sum by the standard Fourier method. Some ingenuity, however, is required to apply the circle method (a variant of Kloosterman) to treat diophantine equations with a relatively small number of variables, an excellent example being the work of Heath-Brown on cubic forms [**HB6**].

It is possible in some circumstances to beat the bound for exponential sums which is derived by the Riemann Hypothesis. This is because the angles of the roots of the L-function themselves vary so that additional cancellation may occur. Deligne and Katz have established such occurrences for families parameterized by points on curves or varieties. In other words, in their cases one is actually considering exponential sums in more variables. However, the cancellation of roots can also occur for families parameterized by points over small irregular sets. More important for analytic number theory is that these sets can be quite general, no structure of a subvariety is needed, but instead a kind of a bilinear form structure would suffice. In practice it is not clear how to work with the roots, so one returns to the corresponding exponential sums where manipulations with the parameters (grouping, gluing, etc.) can be performed properly according to the shape of the involved rational function which is seen with the naked eye. In this process one must not destroy the complete variables since in our mind the corresponding summations are already executed in terms of the roots. Therefore when applying Cauchy's inequality to smooth the one variable composed out of the parameters, we put all the complete variables to the inner summation, say n of them, together with some remaining parameters which where not used in the composition. These inner parameters are critical for enlarging the diagonal. After squaring out we get a complete exponential sum in $2n + 1$ variables which depends on the inner parameters. Except for a few configurations of those, the complete exponential sum satisfies the best possible bound derived from the Riemann Hypothesis, thus saving the factor of square root of the modulus per each variable. Since the number of variables is larger than twice the original, we win the game. The above recipe is somewhat oversimplified, yet it reveals the source of extra saving. One can see how it works in the particular case of [**CI1**]. Speaking of [**CI1**] we would like to add that here the exponential/character sums in several variables emerge after applications of harmonic analysis with respect to the hyperbolic Laplace operator rather than by the traditional Fourier analysis.

The profound theory of Deligne and other geometers is being used in analytic number theory with spectacular effects, yet more ideas need to be invented to fully exploit its potential. Perhaps one should go beyond borrowing estimates and penetrate deeply inside the theory. This is a great subject for future research. P. Michel [**M3**] made the first significant steps (see also [**FK**] and [**KS1**]).

Bibliography

[AS1] A. Adolphson and S. Sperber, *On twisted exponential sums*, Math. Ann. **290** (1991), 713–726.

[AS2] A. Adolphson and S. Sperber, *Character sums in finite fields*, Compositio Math. **52** (1984), 325–354.

[AS3] A. Adolphson and S. Sperber, *Newton polyhedra and the total degree of the L-function associated to an exponential sum*, Invent. math. **88** (1987), 555–569.

[Ahl] L. Ahlfors, *Complex Analysis*, McGraw Hill, 1978.

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

[AL] A.O.L Atkin and J. Lehner, *Hecke operators on* $\Gamma_0(m)$, Math. Ann. **185** (1970), 134–160.

[At] F. V. Atkinson, *The mean value of the zeta-function on the critical line*, Proc. London Math. Soc. (2) **47** (1941), 174–200.

[B] A. Baker, *Transcendental Number Theory*, Cambridge Univ. Press, 1990.

[BH] R. Baker and G. Harman, *The difference between consecutive primes*, Proc. London Math. Soc. **72** (1996), 261–280.

[Ba] W. Banks, *Twisted symmetric-square L-functions and the nonexistence of Siegel zeros on* $GL(3)$, Duke Math. J. **87** (1997), 343–353.

[Ba] M. B. Barban, *The "large sieve" method and its application to number theory*, Uspehi Mat. Nauk **21** (1966), 51–102; English transl. in Russian Math. Surveys **21** (1966), 49–103.

[BG] J. Bernstein and S. Gelbart, *An introduction to the Langlands program*, Birkhaüser, 2003.

[BL] H. Bohr and E. Landau, *Sur les zéros de la fonction* $\zeta(s)$ *de Riemann*, Compte Rendus de l'Acad. des Sciences (Paris) **158** (1914), 106–110.

[Bo1] E. Bombieri, *Maggiorazione del resto nel "Primzahlsatz" col metodo di Erdős–Selberg.*, Ist. Lombardo Accad. Sci. Lett. Rend. A **96** (1962), 343–350.

[Bo2] E. Bombieri, *Le grand crible dans la théorie analytique des nombres*, S.M.F, 1974.

[Bo3] E. Bombieri, *Counting points on curves over finite fields (d'après S. A. Stepanov)*, Lecture Notes in Math. **383** (1974), 234–241.

[Bo4] E. Bombieri, *On exponential sums in finite fields, II*, Invent. math. **47** (1978), 29–39.

[Bo5] E. Bombieri, *On the large sieve*, Mathematika **12** (1965), 201–225.

[BD] E. Bombieri and H. Davenport, *Some inequalities involving trigonometrical polynomials*, Ann. Scuola Norm. Sup. Pisa (3) **23** (1969), 223–241.

[BFI] E. Bombieri, J. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), 203–251.

[BI] E. Bombieri and H. Iwaniec, *Some mean-value theorems for exponential sums*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. **13** (1986), 473–486.

[Bou] J. Bourgain, *Remarks on Montgomery's conjectures on Dirichlet sums*, Lecture Notes in Math. **1469** (1991), 153–165.

[BDCT] C. Breuil, B. Conrad, F. Diamond and R. Taylor, *On the modularity of elliptic curves over* $\mathbb{Q}$*: wild* 3*-adic exercises*, J. Amer. Math. Soc. **14** (2001), 843–939.

[BC] W. E. Briggs and S. Chowla, *On discriminants of binary quadratic forms with a single class in each genus*, Canad. J. Math. **6** (1954), 463–470.

[Bru] R. W. Bruggeman, *Fourier coefficients of cusp forms*, Invent. math. **45** (1978), 1–18.

[Br1] V. Brun, *Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare*, Archiv for Math.og Naturvid. **B34** (1915), no. 8.

[Br2] V. Brun, *Le crible d'Eratosthène et le théorème de Goldbach*, C. R. Acad. Sci. Paris **168** (1919), 544–546.

[Bu] D. Bump, *Automorphic forms and representations*, Cambridge Univ. Press, 1996.

[BDHI] D. Bump, W. Duke, J. Hoffstein and H. Iwaniec, *An estimate for the Hecke eigenvalues of Maass forms*, Internat. Math. Res. Notices **4** (1992), 75–81.

[Bur1] D. A. Burgess, *On character sums and L-series, I*, Proc. London Math. Soc. (3) **12** (1962), 193–206.

[Bur2] D. A. Burgess, *On character sums and L-series, II*, Proc. London Math. Soc. (3) **13** (1963), 524–536.

[BH] C. J. Bushnell and G. Henniart, *An upper bound on conductors for pairs*, J. Number Theory **65** (1997), 183–196.

[Byk] V. A. Bykovsky, *Spectral expansion of certain automorphic functions and its number-theoretical applications*, Proc. Steklov Inst. (LOMI) **134** (1984), 15–33; English transl. in J. Soviet Math. **36** (1987), 8–21.

[Car] F. Carlson, *Über die Nullstellen der Dirichletschen Reihen und der Riemannschen ζ-Funktion*, Arkiv. für Mat. Astr. och Fysik **15** (1920).

[CF] J.W.S Cassels and A. Frölich, *Algebraic Number Theory*, Academic Press, 1990.

[ChI] Chamizo and H. Iwaniec, *On the sphere problem*, Rev. Mat. Iberoamericana **11** (1995), 417–429.

[Ch] J. R. Chen, *On the representation of a larger even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157–176.

[CS] S. Chowla and A. Selberg, *On Epstein's zeta-function*, J. Reine angew. Math. **227** (1967), 86–110.

[Co1] B. Conrey, *Zeros of derivatives of Riemann's ξ-function on the critical line*, J. Number Theory **16** (1983), 49–74.

[Co2] B. Conrey, *More than two fifths of the zeros of the Riemann zeta function are on the critical line*, J. Reine angew. Math. **399** (1989), 1–26.

[CGG] B. Conrey, A. Ghosh and S. M. Gonek, *A note on gaps between zeros of the zeta function*, Bull. London Math. Soc. **16** (1984), 421–424.

[CI1] B. Conrey and H. Iwaniec, *The cubic moment of central values of automorphic L-functions*, Ann. of Math. (2) **151** (2000), 1175–1216.

[CI2] B. Conrey and H. Iwaniec, *Spacing of zeros of Hecke L-functions and the class number problem*, Acta Arithmetica **103** (2002), 259–312.

[CoSo] B. Conrey and K. Soundararajan, *Real zeros of quadratic Dirichlet L-functions*, Invent. math. **150** (2002), 1–44.

[Cor1] J.G. van der Corput, *Zahlentheoretische Abschätzungen*, Math. Ann. **84** (1921), 53–79.

[Cor2] J.G. van der Corput, *Verscharfung der Abschätzungen beim Teilerproblem*, Math. Ann. **87** (1922), 39–65.

[Cor3] J. G. van der Corput, *Sur l'hypothèse de Goldbach pour presque tous les nombres premiers*, Acta Arithmetica **2** (1937), 266–290.

[Cox] D. Cox, *Primes of the form $x^2 + ny^2$*, Wiley, 1989.

[Cra] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Prace Mat.-Fiz. **45** (1937), 51–74.

[Da1] H. Davenport, *On some infinite series involving arithmetical functions. II*, Quart. J. Math. Oxf. **8** (1937), 313–320.

[Da2] H. Davenport, *Analytic methods for Diophantine equations and Diophantine inequalities*, Ann Arbor Publishers, 1963.

[DH1] H. Davenport and H. Halberstam, *The values of a trigonometrical polynomial at well spaced points*, Mathematika **13** (1966), 91–96.

[DH2] H. Davenport and H. Halberstam, *Primes in arithmetic progressions*, Michigan Math. J. **13** (1966), 485–489.

[DHe] H. Davenport and Heilbronn, *On the class-number of binary cubic forms, I, II*, J. London Math. Soc. **26** (1951), 183–192, 192–198.

[De1] P. Deligne, *La conjecture de Weil, I*, Inst. Hautes Études Sci. Publ. Math. **43** (1972), 206–226.

[De2] P. Deligne, *La conjecture de Weil, II*, Inst. Hautes Études Sci. Publ. Math. **52** (1980), 137–252.

[De3] P. Deligne, *Cohomologie étale, SGA* $4\frac{1}{2}$, Lecture Notes. Math. 569, Springer Verlag, 1977.

[DeSe] P. Deligne and J-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. (4) **7** (1974), 507–530.

[DI] J.M. Deshouillers and H. Iwaniec, *Kloosterman sums and Fourier coefficients of cusp forms*, Invent. math. **70** (1982/83), 219–288.

[DS] H. Diamond and J. Steinig, *An elementary proof of the prime number theorem with a remainder term*, Invent. math. **11** (1970), 199–258.

[Dir] P.G. Dirichlet, *Démonstration d'une propriété analogue à la loi de réciprocité qui existe entre deux nombres premiers quelconques*, J. Reine angew. Math. **9** (1832), 379–389.

[Du1] W. Duke, *The dimension of the space of cusp forms of weight one*, Internat. Math. Res. Notices (1995), 99–109.

[Du2] W. Duke, *Hyperbolic distribution problems and half-integral weight Maass forms*, Invent. math. **92** (1988), 73–90.

[Du3] W. Duke, *The critical order of vanishing of automorphic L-functions with large level*, Invent. math. **119** (1995), 165–174.

[Du4] W. Duke, *Some problems in multidimensional analytic number theory*, Acta Arithmetica **52** (1989), 203–228.

[Du5] W. Duke, *Elliptic curves with no exceptional primes*, C. R. Acad. Sci. Paris Sér. I Math. **325** (1997), 813–818.

[DFI1] W. Duke, J. Friedlander and H. Iwaniec, *A quadratic divisor problem*, Invent. math. **115** (1994), 209–217.

[DFI2] W. Duke, J. Friedlander and H. Iwaniec, *Bounds for automorphic L-functions, II*, Invent. math. **115** (1994), 219–239.

[DFI3] W. Duke, J. Friedlander and H. Iwaniec, *The subconvexity problem for Artin L-functions*, Invent. math. **149** (2002), 489–577.

[DFI4] W. Duke, J. Friedlander and H. Iwaniec, *Bilinear forms with Kloosterman fractions*, Invent. math. **128** (1997), 23–43.

[DFI5] W. Duke, J. Friedlander, H. Iwaniec, *Equidistribution of roots of a quadratic congruence to prime moduli*, Ann. of Math. (2) **141** (1995), 423–441.

[DuI1] W. Duke and H. Iwaniec, *Bilinear forms in the Fourier coefficients of half-integral weight cusp forms and sums over primes*, Math. Ann. **286** (1990), 783–802.

[DuI2] W. Duke and H. Iwaniec, *Estimates for coefficients of L-functions, I*, Automorphic forms and analytic number theory (Montreal, PQ, 1989), CRM, 1989, pp. 43–47.

[DuI3] W. Duke and H. Iwaniec, *Estimates for coefficients of L-functions, II*, Proceedings of the Amalfi Conference on Analytic Number Theory, Univ. Salerno, Salerno, 1992, pp. 71–82.

[DuI4] W. Duke and H. Iwaniec, *Estimates for coefficients of L-functions, IV*, Amer. J. Math. **116** (1994), 207–217.

[DK] W. Duke and E. Kowalski, *A problem of Linnik for elliptic curves and mean-value estimates for automorphic representations*, Invent. math. **139** (2000), 1–39.

[D] F. J. Dyson, *Statistical theory of the energy levels of complex systems, I*, J. Mathematical Phys. **3** (1962), 140–156.

[Ell] J. Ellenberg, *Galois representations attached to $\mathbb{Q}$-curves and the generalized Fermat equation $A^4 + B^2 = C^p$*, Amer. J. Math. (to appear).

[El] P.D.T.A. Elliott, *On inequalities of large sieve type*, Acta Arithmetica **18** (1971), 405–422.

[EK] P. Erdős and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math. **62** (1940), 738–742.

[Est] T. Estermann, *On Goldbach's Problem: Proof that Almost All Even Positive Integers are Sums of Two Primes*, Proc. London Math. Soc. (2) **44** (1938), 307–314.

[Fon] J.M. Fontaine, *Il n'y a pas de variétés abéliennes sur* $\mathbb{Z}$, Invent. math. **81** (1985), 515–538.

[FK] É. Fouvry and N. Katz, *A general stratification theorem for exponential sums, and applications*, J. Reine angew. Math. **540** (2001), 115–166.

[FI] É. Fouvry and H. Iwaniec, *Gaussian Primes*, Acta. Arithmetica **79** (1997), 249–287.

[FoM] É. Fouvry and P. Michel, *Sur le changement de signe des sommes de Kloosterman* (preprint).

[Fri] J. Friedlander, *Primes in arithmetic progressions and related topics*, Analytic Number Theory and Diophantine problems, Birkhaüser, 1987, pp. 125–134.

[FG] J. Friedlander and A. Granville, *Limitations to the equi-distribution of primes, III*, Compositio Math. **81** (1992), 19–32.

[FI1] J. Friedlander and H. Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes*, Ann. of Math. (2) **148** (1998), 945–1040.

[FI2] J. Friedlander and H. Iwaniec, *Summation formulae for coefficients of L-functions* (to appear).

[FI3] J. Friedlander and H. Iwaniec, *A Note on Character Sums*, Contemporary Math. **166** (1994), 295–299.

[FI4] J. Friedlander and H. Iwaniec, *Incomplete Kloosterman sums and a divisor problem*, Ann. of Math. **121** (1985), 319–350.

[Ga1] P.X. Gallagher, *A large sieve density estimate near $\sigma = 1$*, Invent. math. **11** (1970), 329–339.

[Ga2] P.X. Gallagher, *Primes in progressions to prime-power modulus*, Invent. math. **16** (1972), 191–201.

[Ga3] P. X. Gallagher, *Bombieri's mean value theorem*, Mathematika **15** (1968), 1–6.

[Ga4] P. X. Gallagher, *The large sieve and probabilistic Galois theory*, Proc. Sympos. Pure Math., Vol. XXIV, Amer. Math. Soc., 1973, pp. 91–101.

[GM] M. L. Gaudin and M. Mehta, *On the density of eigenvalues of a random matrix*, Nuclear Phys. **18** (1960), 420–427.

[Ge] F. Gerth III, *Extension of conjectures of Cohen and Lenstra*, Expositiones Math. **5** (1987), 181–184.

[GeJ] S. Gelbart and H. Jacquet, *A relation between automorphic representations of $GL(2)$ and $GL(3)$*, Ann. Sci. École Norm. Sup. (4) **11** (1978), 471–542.

[GJ] R. Godement and H. Jacquet, *Zeta functions of simple algebras*, Lecture Notes Math. 260, Springer Verlag, 1972.

[Go1] D. Goldfeld, *A simple proof of Siegel's theorem*, Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 1055.

[Go2] D. Goldfeld, *The class number of quadratic fields and the conjectures of Birch and Swinnerton-Dyer*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **3** (1976), 624–663.

[GS] D. Goldfeld and P. Sarnak, *Sums of Kloosterman sums*, Invent. math. **71** (1983), 243–250.

[GHL] D. Goldfeld, J. Hoffstein and D. Lieman, Annals of Math. (2) **140** (1994), 161–181.

[GHB] D. Goldston and D. R. Heath-Brown, *A note on the differences between consecutive primes*, Math. Ann. **266** (1984), 317–320.

[GR] I.S. Gradhsteyn and I.M. Rizhik, *Table of integrals, series and products*, 6th Edition, Academic Press, 2000.

[G1] S. Graham, *An asymptotic estimate related to Selberg's sieve*, J. Number Theory **10** (1978), 83–94.

[G2] S. Graham, *On Linnik's constant*, Acta Arithmetica **39** (1981), 163–179.

[GK] S.W Graham and G. Kolesnik, *van der Corput's method of exponential sums*, Cambridge Univ. Press, 1991.

[GRi] S. W. Graham and C. J. Ringrose, *Lower bounds for least quadratic nonresidues*, Analytic number theory (Allerton Park, IL, 1989), Birkhäuser, 1990, pp. 269–309.

[Gra] A. Granville, *Unexpected irregularities in the distribution of prime numbers*, Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994), Birkhäuser, Basel, 1995, pp. 388–399.

[Gr] G. Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), vol. 43, Springer Verlag, 2001.

[GZ1] B. Gross and D. Zagier, *Heegner points and derivatives of L-series*, Invent. math. **84** (1986), 225–320.

[GZ2] B. Gross and D. Zagier, *Points de Heegner et dérivées de fonctions L*, C. R. Acad. Sci. Paris Sér. I Math. **297** (1983), 85–87.

[HaTu] G. Halász and P.Turán, *On the distribution of roots of Riemann zeta and allied functions, I*, J. Number Theory **1** (1969), 121–137.

[Hal] H. Halberstam, *On the distribution of additive number-theoretic functions, II, III*, J. London Math. Soc. **31** (1956), 1–14, 14–27.

[HaRi] H. Halberstam and H. E. Richert, *Sieve methods*, Academic Press, 1974.

[HaLa] G. H. Hardy and E. Landau, *The lattice points of a circle*, Proc. Royal Soc. A **105** (1924), 244–258.

[HL1] G. H. Hardy and J. E. Littlewood, *Some Problems of 'Partitio Numerorum.' III. On the Expression of a Number as a Sum of Primes.*, Acta Math. **44** (1922), 1–70.

[HL2] G. H. Hardy and J. E. Littlewood, *The zeros of Riemann's zeta function on the critical line*, Math. Z. **10** (1921), 283–317.

[HR] G.H. Hardy and S. Ramanujan, *Asymptotic Formulae in Combinatory Analysis*, Proc. London Math. Soc. **17** (1918), 75–115.

[HW] G.H. Hardy and E.M. Wright, *An introduction to the theory of numbers*, 5th edition, Oxford University Press, 1979.

[HT] M. Harris and R. Taylor, *The geometry and cohomology of some simple Shimura varieties*, Princeton Univ. Press, 2002.

[Ha] R. Hartshorne, *Algebraic Geometry*, Grad. Texts in Math. 52, Springer Verlag, 1977.

[HB1] D.R. Heath-Brown, *A mean value estimate for real character sums*, Acta Arithmetica **72** (1995), 235–275.

[HB2] D.R. Heath-Brown, *An estimate for Heilbronn's exponential sum*, Analytic number theory, Vol. 2 (Allerton Park, IL, 1995), Birkhaüser, 1995, pp. 451–463.

[HB3] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), 1365–1377.

[HB4] D. R. Heath-Brown, *Zero-free regions for Dirichlet L-functions, and the least prime in an arithmetic progression*, Proc. London Math. Soc. (3) **64** (1992), 265–338.

[HB5] D. R. Heath-Brown, *Lattice points in the sphere*, Number theory in progress, Vol. 2, de Gruyter, Berlin, 1999, pp. 883–892.

[HB6] D. R. Heath-Brown, *Cubic forms in ten variables*, Proc. London Math. Soc. (3) **47** (1983), 225–257.

[HBP] D.R. Heath-Brown and S.J. Patterson, *The distribution of Kummer sums at prime arguments*, J. Reine angew. Math. **310** (1979), 111–130.

[Hec1] E. Hecke, *Über eine neue Art von Zetafunktionen*, Math. Zeit. **6** (1920), 11–51.

[Hee] K. Heegner, *Diophantische Analysis und Modulfunktionen*, Math. Z. **56** (1952), 227–253.

[H] D. Hilbert, *Beweis für die Darstellbarkeit der ganzen Zahlen durch eine feste Anzahl n-ter Potenzen (Waringsches Problem)*, Math. Annalen **67** (1909), 281–305.

[Hi1] A. Hildebrand, *An asymptotic formula for the variance of an additive function*, Math. Z. **183** (1983), 145–170.

[Hi2] A. Hildebrand, *On the constant in the Pólya-Vinogradov inequality*, Canad. Math. Bull. **31** (1988), 347–352.

[Hof] J. Hoffstein, *On the Siegel-Tatuzawa theorem*, Acta Arithmetica **38** (1980/81), 167–174.

[HL] J. Hoffstein and P. Lockhart, *Coefficients of Maass forms and the Siegel zero*, Annals of Math. **140** (1994), 161–181.

[Ho] G. Hoheisel, *Primzahlprobleme in der Analysis*, S.-B. Preuss. Akad. Wiss. Phys.-Math. Kl. (1930), 580–588.

[H1] Loo-Keng Hua, *Introduction to number theory*, Springer, 1982.

[H2] Loo Keng Hua, *On Waring's problem*, Quart. J. Math. Oxford **9** (1938), 199–202.

[Hub] H. Huber, *Zur analytischen Theorie hyperbolischen Raumformen und Bewegungsgruppen*, Math. Ann. **138** (1959), 1–26.

[Hu1] M. N. Huxley, *The large sieve inequality for algebraic number fields*, Mathematika **15** (1968), 178–187.

[Hu2] M. N. Huxley, *Large values of Dirichlet polynomials*, Acta Arithmetica **24** (1973), 329–346.

[Hu3] M. N. Huxley, *On the differences between consecutive primes*, Invent. math. **15** (1972), 164–170.

[Hu4] M. N. Huxley, *Area, lattice points, and exponential sums*, The Clarendon Press, 1996.

[HuW] M. N. Huxley and N. Watt, *Exponential sums and the Riemann zeta function*, Proc. London Math. Soc. **57** (1988), 1–24.

[Ing] A. E. Ingham, *On the estimation of $N(\sigma, T)$*, Quart. J. Math. **11** (1940), 291–292.

[IR] K. Ireland and M. Rosen, *A Classical Introduction to Modern Number Theory*, 2nd Edition, Grad. Texts in Math. 84, Springer-Verlag, 1990.

[Iv] A. Ivić, *The Riemann zeta-function, Theory and applications*, Dover Publications, 2003.

[I1] Iwaniec, *Character sums and small eigenvalues for* $\Gamma_0(p)$, Glasgow Math. J. **27** (1985), 99–116.

[I2] H. Iwaniec, *On zeros of Dirichlet's L series*, Invent. math. **23** (1974), 97–104.

[I3] H. Iwaniec, *Spectral theory of automorphic functions and recent developments in analytic number theory*, Proceedings of the ICM Berkeley 1986, Amer. Math. Soc., 1987, pp. 444–456.

[I4] H. Iwaniec, *Topics in Classical Automorphic Forms*, A.M.S, 1997.

[I5] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, 2nd edition, A.M.S and R.M.I, 2002.

[I6] H. Iwaniec, *Fourier coefficients of modular forms of half-integral weight*, Invent. math. **87** (1987), 385–401.

[I7] H. Iwaniec, *The spectral growth of automorphic L-functions*, J. Reine angew. Math. **428** (1992), 139–159.

[I8] H. Iwaniec, *The half-dimensional sieve*, Acta Arithmetica **29** (1976), 69–95.

[I9] H. Iwaniec, *A new form of the error term in the linear sieve*, Acta Arithmetica **37** (1980), 307–320.

[I10] H. Iwaniec, *Almost primes represented by quadratic polynomials*, Invent. math. **47** (1978), 171–188.

[I11] H. Iwaniec, *Small eigenvalues of Laplacian for* $\Gamma_0(N)$, Acta Arithmetica **56** (1990), 65–82.

[I12] H. Iwaniec, *Prime geodesic theorem*, Journal Reine angew. Math. **349** (1984), 136–159.

[I13] H. Iwaniec, *Nonholomorphic modular forms and their applications*, Modular forms (Durham, 1983), Horwood, 1984, pp. 157–196.

[I14] H. Iwaniec, *The lowest eigenvalue for congruence groups*, Topics in geometry, Birkhäuser, 1996, pp. 203–212.

[ILS] H. Iwaniec, W. Luo and P. Sarnak, *Low-lying zeros of families of L-functions*, Inst. Hautes Études Sci. Publ. Math. **91** (2001), 55–131.

[IM] H. Iwaniec and P. Michel, *The second moment of the symmetric square L-functions*, Ann. Acad. Sci. Fenn. Math. **26** (2001), 465–482.

[IS1] H. Iwaniec and P. Sarnak, *Perspectives on the analytic theory of L-functions*, GAFA **Special Volume GAFA2000** (2000), 705–741.

[IS2] H. Iwaniec and P. Sarnak, *The non-vanishing of central values of automorphic L-functions and Landau-Siegel zeros*, Israel J. Math. **120** (2000), 155–177.

[JS] H. Jacquet and J. Shalika, *On Euler products and the classification of automorphic representations, I and II*, Amer. J. Math. **103** (1981), 499–588, 777–815.

[JPS] H. Jacquet, I. Piatetskii-Shapiro and J. Shalika, *Rankin-Selberg convolutions*, Amer. J. Math. **105** (1983), 367–464.

[Ju1] M. Jutila, *Lectures on a method in the theory of exponential sums*, Springer Verlag, 1987.

[Ju2] M. Jutila, *On character sums and class numbers*, J. Number Theory **5** (1973), 203–214.

[Ju3] M. Jutila, *On large values of Dirichlet polynomials*, Topics in number theory (Proc. Colloq., Debrecen, 1974), North-Holland, 1976, pp. 129–140.

[Ju4] M. Jutila, *Zero-density estimates for L-functions*, Acta Arithmetica **32** (1977), 55–62.

[Ju5] M. Jutila, *Statistical Deuring-Heilbronn phenomenon*, Acta Arithmetica **37** (1980), 221–231.

[K1] N. Katz, *Twisted L-functions and monodromy*, Princeton Univ. Press, 2002.

[K2] N. Katz, *Exponential sums over finite fields and differential equations over the complex numbers: some interactions*, Bull. Amer. Math. Soc. (N.S.) **23** (1990), 269–309.

[K3] N. Katz, *Gauss sums, Kloosterman sums and monodromy groups*, Princeton Univ. Press, 1988.

[K4] N. Katz, *Sommes exponentielles*, S.M.F, 1980.

[K5] N. Katz, *Sums of Betti numbers in arbitrary characteristic*, Finite Fields Appl. **7** (2001), 29–44.

[KS1] N. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, A.M.S, 1999.

[KS2] N. Katz and P. Sarnak, *Zeroes of zeta functions and symmetry*, Bull. Amer. Math. Soc. **36** (1999), 1–26.

[Kh] A. Y. Khinchine, *Three pearls of number theory*, Dover Publications, 1998.

[KiS] H. Kim and P. Sarnak, *Refined estimates towards the Ramanujan and Selberg conjectures*, J. Amer. Math. Soc. **16** (2003), 175–181.

[KSh] H. Kim and F. Shahidi, *Cuspidality of symmetric powers with applications*, Duke Math. J. **112** (2002), 177–197.

[Klo] H.D. Kloosterman, *On the representation of numbers in the form* $ax^2+by^2+cz^3+dt^2$, Acta Math. **49** (1926), 407–464.

[Ko1] E. Kowalski, *Analytic problems for elliptic curves* (2001) (preprint).

[Ko2] E. Kowalski, *On the "reducibility" of arctangents of integers*, Amer. Math. Monthly **111** (2004), 351–354.

[KM1] E. Kowalski and P. Michel, *A lower bound for the rank of* $J_0(q)$, Acta Arithmetica **94** (2000), 303–343.

[KM2] E. Kowalski and P. Michel, *The analytic rank of* $J_0(q)$ *and zeros of automorphic L-function*, Duke Math. J. **100** (1999), 503–542.

[KMV1] E. Kowalski, P. Michel and J. VanderKam, *Mollification of the fourth moment of automorphic L-functions and arithmetic applications*, Invent. math. **142** (2000), 95–151.

[KMV2] E. Kowalski, P. Michel and J. VanderKam, *Rankin-Selberg L-functions in the level aspect*, Duke Math. J **114** (2002), 123–191.

[KMV3] E. Kowalski, P. Michel and J. VanderKam, *Non-vanishing of high derivatives of automorphic L-functions at the center of the critical strip*, J. Reine angew. Math. **526** (2000), 1–34.

[Kor] N. M. Korobov, *Estimates of trigonometric sums and their applications*, Uspehi Mat. Nauk. **13** (1958), 185–192.

[Kub1] J. Kubilius, *Probability methods in number theory*, Usp. Mat. Nauk. **68** (1956), 31–66.

[Kub2] J. Kubilius, *Sharpening of the estimate of the second central moment for additive arithmetical functions*, Litovsk. Mat. Sb. **25** (1985), 104–110.

[Kuz] N. V. Kuznetsov, *The Petersson conjecture for cusp forms of weight zero and the Linnik conjecture*, Mat. Sb. (N.S.) **111** (1980), 334–383; Math. USSR-Sb **39** (1981), 299–342.

[LO] J. Lagarias and A. Odlyzko, *Effective versions of the Chebotarev Density Theorem*, Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975), Academic Press, 1977, pp. 409–464.

[La1] E. Landau, *Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate*, Arch. der Math. u. Phys. (3) **13** (1908), 305–312.

[La2] E. Landau, *Bemerkungen zum Heilbronnschen Satz*, Acta Arithmetica **1** (1936), 1–18.

[La3] E. Landau, *Über die Nullstellen der Dirichletschen Reihen und der Riemannschen* ζ*-Funktion*, Arkiv. für Mat. Astr. och Fysik **16** (1921).

[La] S. Lang, *Algebraic Number Theory*, 2nd edition, Grad. Texts in Math. 110, Springer-Verlag, 1994.

[LT] S. Lang and H. Trotter, *Frobenius distribution in* GL_2 *extensions*, Lecture Notes in Math. 504, Springer Verlag, 1976.

[Lau] G. Laumon, *Exponential sums and l-adic cohomology: a survey*, Israel J. Math. **120** (2000), 225–257.

[Lav] A. F. Lavrik, *Approximate functional equations of Dirichlet functions*, Izv. Akad. Nauk SSSR Ser. Mat. **32** (1968), 134–185.

[Leb] N. N. Lebedev, *Special functions and their applications*, Dover Publications, 1972.

[Lev] N. Levinson, *More than one-third of the zeros of the Riemann zetafunction are on* $\sigma = 1/2$, Adv. Math. **13** (1974), 383–436.

[L] W. Li, *L-series of Rankin type and their functional equations*, Math. Ann. **244** (1979), 135–166.

[Li1] Yu. V. Linnik, *The large sieve*, Dokl. Akad. Nauk SSSR **30** (1941), 292–294. (in Russian)

[Li2] Yu. V. Linnik, *The dispersion method in binary additive problems*, AMS, 1963.

[Li3] Yu V. Linnik, *Additive problems and eigenvalues of the modular operators*, Proc. Internat. Congr. Mathematicians (Stockholm, 1962), pp. 270–284.

[Li4] Yu. V. Linnik, *On the least prime in an arithmetic progression, I. The basic theorem*, Rec. Math. [Mat. Sbornik] N.S. **15(57)** (1944), 139–178.

[Li5] Yu. V. Linnik, *On the least prime in an arithmetic progression, II. The Deuring-Heilbronn phenomenon*, Rec. Math. [Mat. Sbornik] N.S. **15(57)** (1944), 347–368.

[Li6] Yu. V. Linnik, *On Dirichlet's L-series and prime-number sums*, Rec. Math. [Mat. Sbornik] N.S. **15(57)** (1944), 3–12.

[Li7] Yu. V. Linnik, *New versions and new uses of the dispersion methods in binary additive problems*, Dokl. Akad. Nauk SSSR **137** (1961), 1299–1302.

[LR] J. H. van Lint and H. -E. Richert, *On primes in arithmetic progressions*, Acta Arithmetica **11** (1965), 209–216.

[Lit] J. E. Littlewood, *On the zeros of the Riemann Zeta-function*, Cambridge Phil. Soc. Proc. **22** (1924), 295–318.

[LRW] J. van de Lune, H.J.J te Riele, D.T. Winter, *On the zeros of the Riemann zeta function in the critical strip, IV*, Math. Comp. **174** (1986), 667–681.

[Luo] W. Luo, *Nonvanishing of L-values and the Weyl law*, Ann. of Math. (2) **154** (2001), 477–502.

[LRS] W. Luo, Z. Rudnick and P. Sarnak, *On Selberg's eigenvalue conjecture*, Geom. Funct. Anal. **5** (1995), 387–401.

[Ma] H. Maass, *Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **121** (1949), 141–183.

[Mai] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), 221–225.

[Mar] G. Margulis, *Discrete Subgroups of Semisimple Lie Groups*, Ergebnisse der Math. und ihrer Grenzgebiete 68, Springer Verlag, 1991.

[MSD] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. math. **25** (1974), 1–61.

[Mes] J-F. Mestre, *Formules explicites et minorations de conducteurs de variétés algébriques*, Compositio Math. **58** (1986), 209–232.

[M1] P. Michel, *The subconvexity problem for Rankin-Selberg L-functions and equidistribution of Heegner points*, Annals. of Math. (to appear).

[M2] P. Michel, *Analytic number theory and families of automorphic L-functions* (to appear).

[M3] P. Michel, *Autour de la conjecture de Sato-Tate pour les sommes de Kloosterman, I*, Invent. Math. **121** (1995), 61–78.

[MvdK] P. Michel and J. VanderKam, *Non-vanishing of high derivatives of Dirichlet L-functions at the central point*, J. Number Theory **81** (2000), 130–148.

[Mi] T. Miyake, *Modular forms*, Springer Verlag, 1989.

[MW] C. Moeglin and J-L. Waldspurger, *Pôles des fonctions L de paires pour $GL(N)$, app. to Le spectre résiduel de $GL(N)$*, Ann. Sci. ENS (4ème série) **22** (1989), 605–674.

[Mol] G. Molteni, *Upper and lower bounds at $s = 1$ for certain Dirichlet series with Euler product*, Duke Math. J. **111** (2002), 133–158.

[Mo1] H.L. Montgomery, *The analytic principle of the large sieve*, Bull. Amer. Math. Soc. **84** (1978), 547–567.

[Mo2] H.L. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Math. 227, Springer Verlag, 1971.

[Mo3] H. Montgomery, *Zeros of L-functions*, Invent. math. **8** (1969), 346–354.

[Mo4] H. L. Montgomery, *The pair correlation of zeros of the zeta function*, Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV), Amer. Math. Soc., 1972, pp. 181–193.

[MV1] H. L. Montgomery and R. C., Vaughan, *The large sieve*, Mathematikia **20** (1973), 119–134.

[MV2] H.L. Montgomery and R.C. Vaughan, *Hilbert's inequality*, J. London Math.Soc. (2) **8** (1974), 73–82.

[MV3] H.L. Montgomery and R.C. Vaughan, *The exceptional set in Goldbach's problem*, Acta Arithmetica **27** (1975), 353–370.

[Mor1] C. Moreno, *Prime number theorems for the coefficients of modular forms*, Bull. Amer. Math. Soc. **78** (1972), 796–798.

[Mor2] C. Moreno, *Algebraic curves over finite fields*, Cambridge Univ. Press, 1991.

[Mor3] C. Moreno, *Analytic proof of the strong multiplicity one theorem*, Amer. J. Math. **107** (1985), 163–206.

[Mot1] Y. Motohashi, *Spectral theory of the Riemann zeta-function*, Camdridge Univ. Press, 1997.

[Mot2] Y. Motohashi, *An induction principle for the generalization of Bombieri's prime number theorem*, Proc. Japan Acad. **52** (1976), 273–275.

[Nak] H. Nakazato, *Heegner points on modular elliptic curves*, Proc. Japan Acad. Ser. A Math. Sci. **72** (1996), 223–225.

[Ne] J. Nekovàr, *On the parity of ranks of Selmer groups, II*, C. R. Acad. Sci. Paris Sér. I Math. **332** (2001), 99–104.

[Od] A. M. Odlyzko, *Some analytic estimates of class numbers and discriminants*, Invent. math. **29** (1975), 275–286.

[Oe] J. Oesterlé, *Nombres de classes des corps quadratiques imaginaires*, Astérisque **121–122** (1985), 309–323.

[Ono] K. Ono, *Nonvanishing of quadratic twists of modular L-functions and applications to elliptic curves*, J. Reine angew. Math. **533** (2001), 81–97.

[PP] A. Perelli and J. Pomykala, *Averages over twisted elliptic L-functions*, Acta Arithmetica **80** (1997), 149–163.

[PSa] Y. Petridis and P. Sarnak, *Quantum unique ergodicity for $SL_2(\mathcal{O})\backslash\mathbb{H}_3$ and estimates for L-functions*, Journal of Evolution Equations **1** (2001), 277–290.

[Ph] E. Phillips, *The zeta-function of Riemann; further developments of van der Corput's method*, Quart. J. Math. **4** (1933), 209–225.

[PS] R. Phillips and P. Sarnak, *On cusp forms for co-finite subgroups of $PSL(2,\mathbb{R})$*, Invent. math. **80** (1985), 339–364.

[Pi] N. Pitt, *On shifted convolution of $\zeta^3(s)$ with automorphic L-functions*, Duke Math. J. **77** (1995), 383–406.

[Poi] G. Poitou, *Sur les petits discriminants*, Séminaire Delange-Pisot-Poitou, 18e année (1976-77).

[Pol] G. Pólya, *Über die Verteilung der quadratischen Reste und Nichtreste*, Nachr. Königl. Gesell. Wissensch. Göttingen, Math.-phys. Klasse (1918), 21–29.

[Pos] A. G. Postnikov, *On Dirichlet L-series with the character modulus equal to the power of a prime number*, J. Indian Math. Soc. (N.S.) **20** (1956), 217–226.

[PR] A. G. Postnikov and N. P. Romanov, *A simplification of A. Selberg's elementary proof of the asymptotic law of distribution of prime numbers*, Uspehi Mat. Nauk (N.S.) **10** (1955), 75–87.

[R] H. Rademacher, *On the Partition Function $p(n)$*, Proc. London Math. Soc. **43** (1937), 241–254.

[Ra] D. Ramakrishnan, *Modularity of the Rankin-Selberg L-series, and multiplicity one for $SL(2)$*, Annals of Math. (2) **152** (2000), 45–111.

[Ra1] R. A. Rankin, *Van der Corput's method and the theory of exponent pairs*, Quart. J. Math. (2) **6** (1955), 147–153.

[Ra2] R. A. Rankin, *The difference between consecutive prime numbers, V*, Proc. Edinburgh Math. Soc. (2) **13** (1962/1963), 331–332.

[Ra3] R. A. Rankin, *Contributions to the theory of Ramanujan's τ function and similar arithmetical functions, II*, Proc. Camb. Phil. Soc. **35** (1939), 351–372.

[Re] A. Renyi, *On the representation of an even number as the sum of a single prime and single almost-prime number*, Izvestiya Akad. Nauk SSSR. Ser. Mat. **12** (1948), 57–78.

[Rey] É. Reyssat, *Quelques aspects des surfaces de Riemann*, Progress in Math. 77, Birkhaüser, 1989.

[Rie] B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsber. Berlin. Akad. (1859), 671–680.

[RV] F. Rodriguez-Villegas, *Square root formulas for central values of Hecke L-series, II*, Duke Math. J. **72** (1993), 431–440.

[Rot] K.F. Roth, *On the large sieves of Linnik and Rényi*, Mathematika **12** (1965), 1–9.

[Roy] E. Royer, *Statistique de la variable aléatoire $L(Sym^2 f, 1)$*, Math. Ann. **321** (2001), 667–687.

[Ru] W. Rudin, *Real and complex analysis*, McGraw-Hill, 1987.

[RS] Z. Rudnick and P. Sarnak, *Zeros of principal L-functions and random matrix theory*, Duke Math. J. **81** (1996), 269–322.

[Sal] H. Salié, *Über die Kloostermanschen Summen $S(u, v, q)$*, Math. Z. **34** (1931), 91–109.

[Sa1] P. Sarnak, *Estimates for Rankin-Selberg L-Functions and Quantum Unique Ergodicity*, J. Funct. Anal. **184** (2001), 419–453.

[Sa2] P. Sarnak, *Class Numbers of indefinite binary quadratic forms*, Journal of Number Theory **15** (1982), 229–247.

[Sa3] P. Sarnak, *Some applications of modular forms*, Cambridge Univ. Press, 1990.

[Sa4] P. Sarnak, *Arithmetic quantum chaos*, Bar-Ilan Univ., 1995.

[Sch] W. Schmidt, *Equations over finite fields. An elementary approach*, Lecture Notes in Math. 536, Springer Verlag, 1976.

[ST] D. B. Sears and E.C. Titchmarsh, *Some eigenfunction formulae*, Quart. J. Math. Oxford **1** (1950), 165–175.

[S1] A. Selberg, *The general sieve method and its place in prime number theory*, Proc. ICM, vol. 1, Cambridge, MA., 1950, pp. 286–292.

[S2] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Sympos. Pure Math., Vol. VIII, Amer. Math. Soc., 1965, pp. 1–15.

[S3] A. Selberg, *Lectures on sieves*, Collected Papers Vol. II, Springer Verlag, 1991, pp. 66–247.

[S4] A. Selberg, *On the zeros of Riemann's zeta-function*, Skr. Norske Vid. Akad. Oslo I (1942).

[S5] A. Selberg, *Bemerkungen über eine Dirichletsche Reihe, die mit der Theorie der Modulformen nahe verbunden ist*, Arch. Math. Naturvid. **43** (1940), 47–50.

[S6] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) **20** (1956), 47–87.

[S7] A. Selberg, *On the estimation of Fourier coefficients of modular forms*, Proc. Symp. Pure Math. 8, 1965, pp. 1–8.

[S8] A. Selberg, *On the zeros of the zeta function of Riemann*, Der Kong. Norske Vidensk. Selsk. Forhand. **15** (1942), 59–62.

[Se1] J-P. Serre, *Cours d'arithmétique*, 2nd edition, P.U.F, 1977.

[Se2] J-P. Serre, *Modular forms of weight one and Galois representations*, Algebraic number fields: L-functions and Galois properties (Proc. Sympos., Univ. Durham, Durham, 1975), Academic Press, 1977, pp. 193–268.

[Se3] J-P. Serre, *Congruences et formes modulaires (d'après H. P. F. Swinnerton-Dyer)*, Séminaire Bourbaki, 24e année (1971/1972), Exp. No. 416, Lecture Notes in Math. 317, Springer Verlag, 1973, pp. 319–338.

[Se4] J-P. Serre, *Corps locaux*, Hermann, 1968.

[Se5] J-P. Serre, *Représentations linéaires des corps finis*, Hermann, 1971.

[Se6] J-P. Serre, *Quelques applications du théorème de densité de Chebotarev*, Publ. Math. IHES **54** (1981), 123–201.

[Se7] J-P. Serre, *Minorations de discriminants*, Oeuvres, Vol. III, Springer-Verlag, 1986, pp. 240–243.

[Se8] J-P. Serre, *Letter to J.M. Deshouillers*.

[Shah] F. Shahidi, *Symmetric power L-functions for $GL(2)$*, Elliptic curves and related topics, CRM Proc. Lecture Notes 4, Amer. Math. Soc., 1994, pp. 159–182.

[Sha] D. Shanks, *Class number, a theory of factorization, and genera*, 1969 Number Theory Institute (Proc. Sympos. Pure Math., Vol. XX), Amer. Math. Soc., 1971, pp. 415–440.

[Sh1] G. Shimura, *On modular forms of half-integral weight*, Annals of Math. **97** (1973), 440–481.

[Sh2] G. Shimura, *On the holomorphy of certain Dirichlet series*, Proc. London Math. Soc. (3) **31** (1975), 79–98.

[Sh3] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Princeton Univ. Press, 1971.

[Sie1] C. L. Siegel, *Über die Classenzahl quadratischer Zahlkörper*, Acta Arithmetica **1** (1936), 83–86.

[Sie2] C.L. Siegel, *On the theory of indefinite quadratic forms*, Ann. of Math. **45** (1944), 577–622.

[Sie3] C. L. Siegel, *Lectures on quadratic forms*, Tata Institute, 1967.

[Sil] J. Silverman, *The arithmetic of elliptic curves*, Grad. Texts in Math. 106, Springer-Verlag, 1986.

[Sou] K. Soundararajan, *Nonvanishing of quadratic Dirichlet L-functions at* $s = \frac{1}{2}$, Ann. of Math. (2) **152** (2000), 447–488.

[St1] H. Stark, *A complete determination of the complex quadratic fields with class-number one*, Michigan Math. J. **14** (1967), 1–27.

[St2] H. Stark, *A transcendence theorem for class-number problems. II*, Annals of Math. (2) **96** (1972), 174–209.

[St3] H. Stark, *Some effective cases of the Brauer-Siegel theorem*, Invent. math. **23** (1974), 135–152.

[Ste] S. A. Stepanov, *The number of points of a hyperelliptic curve over a finite prime field*, Izv. Akad. Nauk SSSR Ser. Mat. **33** (1969), 1171–1181.

[Ta1] J. Tate, *Fourier analysis in number fields and Hecke's zeta functions*, Algebraic Number Theory, Academic Press, 1990, pp. 305–347.

[Tat] J. Tate, *Number theoretic preliminaries*, Proceedings of Symposia in Pure Math. 33, vol 2, A.M.S, 1979, pp. 3–26.

[TW] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. of Math. (2) **141** (1995), 553–572.

[Tchu] N. G. Tchudakov, *Sur le problème de Goldbach*, C. R. (Dokl.) Acad. Sci. URSS, n. Ser. **17** (1937), 335–338.

[Th] A. Thue, *Über Annäherungswerte algebraischer Zahlen*, J. Reine angew. Math. **135** (1909), 284–305.

[T1] E.C. Titchmarsh, *The theory of functions*, 2nd Edition, Oxford Univ. Press, 1939.

[T2] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd edition, Oxford Univ. Press, 1986.

[T3] E. C. Titchmarsh, *Eigenfunction expansions associated with second-order differential equations*, Clarendon Press, 1962.

[Tot] A. Toth, *Roots of quadratic congruences*, Internat. Math. Res. Notices (2000), 719–739.

[Tu1] P. Turán, *Über einige Verallgemeinerungen eines Satzes von Hardy und Ramanujan*, J. Lond. Math. Soc. **11** (1936), 125–133.

[Tu2] P. Turán, *Über die Primzahlen der arithmetischen Progression*, Acta Litt. Sci. Szeged **8** (1937), 226–235.

[vdK] J. VanderKam, *The rank of quotients of* $J_0(N)$, Duke Math. J. **97** (1999), 545–577.

[VdP] M. van der Put, *Grothendieck's conjecture for the Risch equation* $y' = ay + b$, Indag. Mathem. N.S. **12** (2001), 113–124.

[Va] R. C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), A981–A983.

[V1] I. M. Vinogradov, *A new estimate for* $\zeta(1 + it)$, Izv. Akad. Nauk SSSR, Ser. Mat. **22** (1958), 161–164.

[V2] I. M. Vinogradov, *On Weyl's sums*, Mat. Sbornik **42** (1935), 521–530.

[V3] I. M. Vinogradov, *A new method of estimation of trigonometrical sums*, Mat. Sbornik (1) **43** (1936), 175–188.

[V4] I. M. Vinogradov, Perm. Univ. Fiz.-Mat. ob.-vo Zh **1** (1918), 18–24.

[V5] I. M.Vinogradov, *Some theorems concerning the theory of primes*, Math. Sb. 2 **44** (1937), 179–195.

[V6] I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, Dokl. Akad. Nauk SSSR **15** (1937), 291–294.

[Vi] A. I. Vinogradov, *The density hypothesis for Dirichet L-series*, Izv. Akad. Nauk SSSR Ser. Mat. **29** (1965), 903–934.

[Vor] G. Voronoi, *Sur une fonction transcendante et ses applications à la sommation de quelques séries*, Ann. Sci. École Norm. Sup. (3) **21** (1904), 207–267, 459–533..

[Was] L. Washington, *Introduction to cyclotomic fields*, Grad. Texts in Math. 83, 2nd edition, Springer Verlag, 1997.

[Wa] T. Watson, *Rankin triple products and quantum chaos, PhD thesis, Princeton University*, 2001.

[We1] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U. S. A. **34** (1948), 204–207.

[We2] A. Weil, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **168** (1967), 149–156.

[W1] H. Weyl, *Über die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), 313–352.

[W2] H. Weyl, *Zur Abschätzung von* $\zeta(1+ti)$, Math. Zeit. **10** (1921), 88–101.

[W] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Ann. of Math. (2) **141** (1995), 443–551.

[Wi1] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplikative Funktionen, II*, Acta Math. Acad. Sci. Hungar. **18** (1967), 411–467.

[Wi2] E. Wirsing, *Elementare Beweise des Primzahlsatzes mit Restglied, II*, J. Reine angew. Math. **214/215** (1964), 1–18.

[Wi3] E. Wirsing, *Growth and differences of additive arithmetic functions*, Topics in classical number theory, Vol. I, II (Budapest, 1981), Colloq. Math. Soc. János Bolyai, 34, North-Holland, Amsterdam, 1984, pp. 1651–1661.