WEIL'S BOUND FOR KLOOSTERMAN SUMS

GERGELY HARCOS

1. Introduction

The aim of these notes is to give a concise but self-contained proof of the following celebrated theorem due to Weil [5].

Theorem 1 (Weil)**.** *Let $p > 2$ be a prime number. Let a and b be integers coprime to p. Then the Kloosterman sum*

$$S(a,b;p) := \sum_{t=1}^{p-1} e_p(at + b\bar{t})$$

has absolute value at most $2\sqrt{p}$.

Here $e_p(x)$ abbreviates $\exp(2\pi ix/p)$, and $\bar{t}$ is a multiplicative inverse of t modulo p. We denote by $\mathbb{F}_p$ the p-element field, and we identify its elements with the residue classes modulo p. Hence $e_p(x)$ is a nontrivial additive character of $\mathbb{F}_p$, and we can write

$$S(a,b;p) = \sum_{t\in\mathbb{F}_p^\times} e_p(at + bt^{-1}). \tag{1}$$

We fix p, a, b for the rest of the notes, except that in the next section p is an arbitrary prime.

Our exposition is largely based on Iwaniec–Kowalski [2, Chapter 11], but we try to give more detail at certain points and keep the algebraic prerequisites to a minimum. A rough outline of the proof is as follows. Along with $S(a,b;p)$, we consider all the Kloosterman sums $S(ma,mb;p)$ with $1 \leqslant m \leqslant p-1$, and we write them as

$$S(ma,mb;p) = -\alpha_m - \beta_m$$

with complex numbers α_m and β_m such that $\alpha_m\beta_m = p$. That is, we have a decomposition of polynomials in $\mathbb{C}[T]$,

$$1 + S(ma,mb;p)T + pT^2 = (1-\alpha_m T)(1-\beta_m T). \tag{2}$$

It turns out that the power sums of the α_m's and β_m's have a geometric meaning, namely

$$p^n - 1 - \sum_{m=1}^{p-1}(\alpha_m^n + \beta_m^n) = |\{(x,y)\in\mathbb{F}_{p^n}\times\mathbb{F}_{p^n} :\ y^2 = (x^p - x)^2 - 4ab\}|, \tag{3}$$

where $\mathbb{F}_{p^n}$ denotes the field of p^n elements. Weil showed [6, p. 70] that the right hand side can be approximated as $p^n + O_p(p^{n/2})$, hence for any integer $n \geqslant 1$ we have

$$\sum_{m=1}^{p-1}(\alpha_m^n + \beta_m^n) \ll_p p^{n/2}.$$

It is straightforward to deduce from here that each α_m and β_m has absolute value $\sqrt{p}$, and Theorem 1 follows upon noting $|S(a,b;p)| \leqslant |\alpha_1| + |\beta_1|$.

2. Background on finite fields

Lemma 1. *Let F be a field, and let $k(X) \in F[X]$ be an irreducible polynomial. Then there is a field G containing F such that k has a root in G.*

Proof. It suffices to construct a field G such that F embeds into G, and k has a root in G. The residue classes in $F[X]$ modulo $k(X)$ form a ring $G := F[X]/(k(X))$. We claim that G is a field satisfying the requirements. Clearly, the inclusion $F \subset F[X]$ induces an embedding $F \hookrightarrow G$. If $\xi \in G$ denotes the residue class of X modulo $k(X)$, then $k(\xi) \in G$ is the residue class of $k(X)$ modulo $k(X)$, i.e. $k(\xi) = 0$. Now let $a(X)$ mod $k(X)$ be any nonzero residue class in G, i.e. $a(X) \in F[X]$ is not divisible by $k(X)$. Using the Euclidean algorithm in $F[X]$, we can find polynomials $b(X), l(X) \in F[X]$ such that $a(X)b(X) - k(X)l(X) = 1$. More precisely, the Euclidean algorithm finds a nonzero polynomial of the form $a(X)b(X) - k(X)l(X)$ whose divisors are the common divisors of $a(X)$ and $k(X)$. As $k(X)$ is irreducible, and $a(X)$ is not divisible by $k(X)$, the polynomial $a(X)b(X) - k(X)l(X)$ is a nonzero constant. Dividing $b(X)$ and $l(X)$ by this nonzero constant, we can achieve $a(X)b(X) - k(X)l(X) = 1$. This means that $b(X)$ mod $k(X)$ is a multiplicative inverse of $a(X)$ mod $k(X)$, hence G is a field. □

Definition 1. If $F \subset G$ are fields and $\xi \in G$, then $F(\xi)$ denotes the smallest subfield of G containing F and ξ. The element $\xi \in G$ is called *algebraic* over F if $k(\xi) = 0$ for some non-constant $k(X) \in F[X]$. In this case the unique monic polynomial $k(X) \in F[X]$ of smallest positive degree such that $k(\xi) = 0$ is called the *minimal polynomial* of ξ over F.

Lemma 2. *Let $F \subset G$ be any fields. Let $\xi \in G$ be algebraic over F with minimal polynomial $k(X) \in F[X]$. Then $k(X)$ is irreducible in $F[X]$, and $F(\xi)$ is isomorphic to $F[X]/(k(X))$.*

Proof. If $k(X)$ is reducible in $F[X]$, then it factors into smaller degree monic polynomials $k(X) = u(X)v(X)$. As $k(\xi) = 0$, we have $u(\xi) = 0$ or $v(\xi) = 0$, a contradiction. So $k(X)$ is irreducible in $F[X]$. Moreover, using the Euclidean algorithm in $F[X]$, we see that a polynomial $a(X) \in F[X]$ satisfies $a(\xi) = 0$ if and only if $a(X)$ is divisible by $k(X)$ in $F[X]$. Consider now $F[\xi]$, the smallest subring of G containing F and ξ. Consider also the map $f : F[X]/(k(X)) \to F[\xi]$ assigning to any residue class $a(X)$ mod $k(X)$ the element $a(\xi) \in F[\xi]$. It is straightforward to verify that f is a ring isomorphism, hence $F[\xi]$ is isomorphic to $F[X]/(k(X))$. The latter ring is actually a field (cf. proof of Lemma 1), hence $F[\xi]$ is a field. It follows that $F(\xi) = F[\xi]$, and so $F(\xi)$ is isomorphic to $F[X]/(k(X))$. □

Lemma 3. *Let $F \subset G$ be any fields. If F is a finite field of cardinality m, then*

$$F = \{x \in G : \ x^m = x\}.$$

Proof. If $x \in F^\times$, then $x^{m-1} = 1$, because $F^\times$ is a finite group of order $m-1$. Hence for any $x \in F$ we have $x^m = x$, i.e. $F \subset \{x \in G : \ x^m = x\}$. We must have equality here, because the left hand side has cardinality m, and the right hand side has cardinality at most m. □

Lemma 4. *Let F be a field, and let H be a finite subgroup of the multiplicative group $F^\times$. Then H is cyclic.*

Proof. Let $x \in H$ and $y \in H$ be any two group elements of (multiplicative) orders r and s, respectively. We claim that H contains an element of order $[r,s]$. To see this, decompose $[r,s]$ as ab with suitable $a \mid r$ and $b \mid s$ such that $(a,b) = 1$, and consider $z := x^{r/a}y^{s/b} \in H$. The order t of z clearly divides ab, because $z^{ab} = x^{rb}y^{sa} = 1$. On the other hand, $z^t = 1$ implies $z^{at} = 1$ and $z^{bt} = 1$, whence $y^{ats/b} = 1$ and $x^{btr/a} = 1$. This is only possible if $b \mid at$ and $a \mid bt$, i.e. $ab \mid t$. Hence $t = ab = [r,s]$ as claimed. Now pick $x \in H$ so that its order r is maximal. Then any $y \in H$ has order $s \mid r$, because $[r,s] \leqslant r$ implies $s \mid r$. Fixing y and s, we observe that in F the equation $t^s = 1$ has at most s solutions, and y is one of them. On the other hand, $x^{kr/s}$ $(1 \leqslant k \leqslant s)$ are s distinct elements satisfying $t^s = 1$, hence y must be one these elements. That is, x generates H, and we are done. □

Lemma 5. *Let $F \subset G$ be any finite fields. Then there exists $\xi \in G$ such that $F(\xi) = G$.*

Proof. By Lemma 4, the multiplicative group $G^\times$ is generated by some $\xi \in G$. Then $F(\xi)$ clearly contains $G^\times$, hence $F(\xi) = G$. □

Theorem 2. *The cardinality of any finite field is a prime power p^n. Conversely, for any prime power p^n, there is a finite field of cardinality p^n, and it is unique up to isomorphism.*

Proof. Let F be any finite field. The elements 1, $1+1$, $1+1+1$, etc. in F cannot all be distinct. Hence, after subtraction, we see that in F we have $m \cdot 1 = 0$ for some positive integer m. If m is minimal with this property, then m is prime. Indeed, if $m = kl$ with $0 < k, l < m$, then $(k \cdot 1)(l \cdot 1) = m \cdot 1 = 0$, hence $k \cdot 1 = 0$ or $l \cdot 1 = 0$, a contradiction. So $m = p$ is prime, and we can embed $\mathbb{F}_p$ into F by mapping a residue class t mod p in $\mathbb{F}_p$ to $t \cdot 1 \in F$. We call $\mathbb{F}_p$ the *prime field* of F. In particular, F is a vector space over $\mathbb{F}_p$ of some finite dimension n, hence $|F| = p^n$ is a prime power.

Conversely, let p^n be any prime power. We construct a field F_n of cardinality p^n, and in the next paragraph we show that any field of cardinality p^n is isomorphic to F_n. There is a field K containing $\mathbb{F}_p$ such that any polynomial in $\mathbb{F}_p[X]$ decomposes into linear factors over K. To see this, enumerate the polynomials in $\mathbb{F}_p[X]$, and use Lemma 1 recursively to construct a chain of fields

$$\mathbb{F}_p = K_0 \subset K_1 \subset K_2 \subset \dots$$

such that in K_m the m-th polynomial decomposes into linear factors, and then define K as the union of these fields. Now we put

$$F_n := \{x \in K : \ x^{p^n} = x\}, \tag{4}$$

and we claim that F_n is a p^n-element subfield of K containing $\mathbb{F}_p$. On the one hand, in $K[X]$ we have a decomposition

$$X^{p^n} - X = \prod_{i=1}^{p^n} (X - t_i), \tag{5}$$

and F_n is the set of roots $\{t_i : \ 1 \leqslant i \leqslant p^n\}$. The roots are distinct, because the formal derivative of the left hand side, $(X^{p^n} - X)' = p^n X^{p^n - 1} - 1 = -1$, has no root. This proves that $|F_n| = p^n$. On the other hand, F_n is the set of fixed points of σ^n, where

$$\sigma : x \mapsto x^p \tag{6}$$

denotes the *Frobenius map* on K. Clearly, $\sigma(0) = 0$, $\sigma(1) = 1$, $\sigma(xy) = \sigma(x)\sigma(y)$. Moreover, by the binomial theorem, $\sigma(x+y) = \sigma(x) + \sigma(y)$. Therefore, σ is a field endomorphism of K fixing $\mathbb{F}_p$ pointwise, and the same is true of σ^n. Hence F_n is a subfield of K containing $\mathbb{F}_p$.

Let F be any field of cardinality p^n. Then the prime field of F must be $\mathbb{F}_p$, hence without loss of generality, F contains $\mathbb{F}_p$. By Lemma 5, there exists $\xi \in F$ such that $\mathbb{F}_p(\xi) = F$. Let $k(X) \in \mathbb{F}_p[X]$ be the minimal polynomial of ξ over $\mathbb{F}_p$. Then $k(X)$ has a root α in K. The minimal polynomial of α over $\mathbb{F}_p$ divides $k(X)$ in $\mathbb{F}_p[X]$ (cf. proof of Lemma 2), therefore it equals $k(X)$ by the irreducibility of $k(X)$. It follows that $\mathbb{F}_p(\xi)$ is isomorphic to the subfield $\mathbb{F}_p(\alpha) \subset K$, because both fields are isomorphic to $\mathbb{F}_p[X]/(k(X))$ by Lemma 2. In particular, $\mathbb{F}_p(\alpha)$ has cardinality p^n, hence it equals F_n by Lemma 3. In the end, we see that F is isomorphic to F_n, namely $F = \mathbb{F}_p(\xi) \cong \mathbb{F}_p[X]/(k(X)) \cong \mathbb{F}_p(\alpha) = F_n$. □

Definition 2. We identify $\mathbb{F}_{p^n}$ with F_n defined by (4), and we regard their union

$$\overline{\mathbb{F}_p} := \bigcup_{n=1}^{\infty} \mathbb{F}_{p^n}.$$

Corollary 1. *The fields $\mathbb{F}_{p^n}$ are precisely the finite subfields of $\overline{\mathbb{F}_p}$. Moreover, $\mathbb{F}_{p^m} \subset \mathbb{F}_{p^n}$ if and only $m \mid n$.*

Proof. By definition, $\mathbb{F}_{p^n} \subset \overline{\mathbb{F}_p}$. Conversely, let F be a finite subfield of $\overline{\mathbb{F}_p}$. Then, F contains $\mathbb{F}_p$, hence F is a vector space over $\mathbb{F}_p$ of some finite dimension n. It follows that $|F| = p^n$, therefore $F = \mathbb{F}_{p^n}$ by Lemma 3 and (4). Assume now that $\mathbb{F}_{p^m} \subset \mathbb{F}_{p^n}$. Then, $\mathbb{F}_{p^n}$ is a vector space over $\mathbb{F}_{p^m}$ of some finite dimension k, hence $p^n = \mathbb{F}_{p^n} = |\mathbb{F}_{p^m}|^k = p^{mk}$. That is, $n = mk$, i.e. $m \mid n$. Conversely, if $m \mid n$, then (4) readily implies that $\mathbb{F}_{p^m} \subset \mathbb{F}_{p^n}$. □

In particular, any $\alpha \in \overline{\mathbb{F}_p}$ generates some finite field $\mathbb{F}_p(\alpha) = \mathbb{F}_{p^d}$, hence by (4) and (6) we see that $\overline{\mathbb{F}_p}$ is a disjoint union of *Frobenius orbits* of the form $\{\alpha, \sigma(\alpha), \dots, \sigma^{d-1}(\alpha)\}$. In fact, for a given integer $n \geqslant 1$, the orbits of size $d \mid n$ partition $\mathbb{F}_{p^n}$. The next result describes these orbits in more detail and shows that $\overline{\mathbb{F}_p}$ is an algebraic closure of $\mathbb{F}_p$.

Theorem 3. *A Frobenius orbit of size d in $\overline{\mathbb{F}_p}$ is the set of roots of an irreducible monic polynomial of degree d in $\mathbb{F}_p[X]$, and vice versa.*

Proof. The proof relies on the fact that $\mathbb{F}_p$ is the set of fixed points of σ in $\overline{\mathbb{F}_p}$. Let $\{\alpha, \sigma(\alpha), \dots, \sigma^{d-1}(\alpha)\} \subset \overline{\mathbb{F}_p}$ be a Frobenius orbit of size d, i.e. $\sigma^d(\alpha) = \alpha$ and the listed elements are distinct. Then the monic polynomial

$$k(X) := \prod_{i=0}^{d-1} (X - \sigma^i(\alpha))$$

lies in $\mathbb{F}_p[X]$, because σ permutes the roots and therefore fixes the coefficients of $k(X)$. Moreover, $\{\alpha, \sigma(\alpha), \dots, \sigma^{d-1}(\alpha)\}$ is not the disjoint union of two non-empty σ-invariant subsets, hence $k(X)$ is irreducible in $\mathbb{F}_p[X]$. Conversely, let $k(X) \in \mathbb{F}_p[X]$ be an irreducible monic polynomial of degree d. Then, as we have seen in the proof of Theorem 2, $k(X)$ has a root $\alpha \in \mathbb{F}_{p^d}$, and in fact $\mathbb{F}_p(\alpha) = \mathbb{F}_{p^d}$. Hence $\{\alpha, \sigma(\alpha), \dots, \sigma^{d-1}(\alpha)\}$ is a Frobenius orbit of size d in $\overline{\mathbb{F}_p}$, i.e. $\sigma^d(\alpha) = \alpha$ and the listed elements are distinct. Each element $\sigma^i(\alpha)$ is a root of $k(X)$, because $k(\sigma^i(\alpha)) = \sigma^i(k(\alpha)) = \sigma^i(0) = 0$, therefore the orbit is the set of roots of $k(X)$. □

Corollary 2 (Gauss)**.** *For any integer $n \geqslant 1$, we have the following identity in $\mathbb{F}_p[X]$:*

$$X^{p^n} - X = \prod_{d \mid n} \prod_{\substack{k \text{ irred. monic} \\ \deg(k) = d}} k(X),$$

where the inner product runs through the irreducible monic polynomials of degree d in $\mathbb{F}_p[X]$.

Proof. We have seen in the proof of Theorem 2 that over $\mathbb{F}_{p^n}$ the left hand side decomposes into distinct linear factors as (cf. (5))

$$X^{p^n} - X = \prod_{t \in \mathbb{F}_{p^n}} (X - t).$$

The field $\mathbb{F}_{p^n}$ is a disjoint union of the Frobenius orbits of size $d \mid n$, hence the stated identity follows immediately from Theorem 3. □

Definition 3. The n-trace of an element $\alpha \in \mathbb{F}_{p^n}$ is given by

$$\mathrm{Tr}_n(\alpha) := \sum_{i=0}^{n-1} \sigma^i(\alpha).$$

Theorem 4. *The n-trace is an $\mathbb{F}_p$-linear surjection $\mathrm{Tr}_n : \mathbb{F}_{p^n} \to \mathbb{F}_p$. Moreover, for any $y \in \mathbb{F}_{p^n}$, we have*

$$|\{x \in \mathbb{F}_{p^n} : x^p - x = y\}| = \begin{cases} 0, & \mathrm{Tr}_n(y) \neq 0; \\ p, & \mathrm{Tr}_n(y) = 0. \end{cases} \tag{7}$$

Proof. For any $\alpha \in \mathbb{F}_{p^n}$, we have $\sigma^n(\alpha) = \alpha$, hence

$$\sigma(\mathrm{Tr}_n(\alpha)) = \sum_{i=0}^{n-1} \sigma^{i+1}(\alpha) = \mathrm{Tr}_n(\alpha).$$

That is, $\mathrm{Tr}_n(\alpha) \in \mathbb{F}_p$. In addition, the map $\mathrm{Tr}_n : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is $\mathbb{F}_p$-linear, because σ (hence also σ^i) is $\mathbb{F}_p$-linear. Consider now the $\mathbb{F}_p$-linear map $\delta : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ given by

$$\delta(x) := x^p - x = \sigma(x) - x.$$

The kernel of δ equals $\mathbb{F}_p$, hence δ is a p-to-1 map with an image of size $|\operatorname{im}\delta| = p^{n-1}$. In addition, $\operatorname{im}\delta \subset \ker \mathrm{Tr}_n$, because for any $x \in \mathbb{F}_{p^n}$ we have

$$\mathrm{Tr}_n(\delta(x)) = \mathrm{Tr}_n(\sigma(x) - x) = \sum_{i=0}^{n-1} (\sigma^{i+1}(x) - \sigma^i(x)) = \sigma^n(x) - x = 0.$$

However, $\mathrm{Tr}_n : \mathbb{F}_{p^n} \to \mathbb{F}_p$ is also a polynomial function of degree p^{n-1} by definition, hence it cannot vanish at more than p^{n-1} points. It follows that $\operatorname{im}\delta = \ker \mathrm{Tr}_n$. This verifies (7) and the surjectivity of Tr_n as well, because $|\operatorname{im}\mathrm{Tr}_n| = p^n/|\ker \mathrm{Tr}_n| = p$. □

Remark 1. Theorem 4 and its proof can be summarized by saying that the following sequence of $\mathbb{F}_p$-linear maps is exact:

$$0 \longrightarrow \mathbb{F}_p \xrightarrow{\mathrm{id}} \mathbb{F}_{p^n} \xrightarrow{\delta} \mathbb{F}_{p^n} \xrightarrow{\mathrm{Tr}_n} \mathbb{F}_p \longrightarrow 0.$$

3. L-FUNCTIONS

In this section we prove the identity (3) with the help of L-functions. Recall that the parameters $p, a, b \in \mathbb{Z}$ of Theorem 1 are fixed, and the numbers $\alpha_m, \beta_m \in \mathbb{C}$ $(1 \leqslant m \leqslant p-1)$ satisfy (2).

The ring of polynomials $\mathbb{F}_p[X]$ bears a close similarity to the ring of integers $\mathbb{Z}$. We define a completely multiplicative function $\eta : \mathbb{F}_p[X] \to \mathbb{C}$ that is analogous to a Dirichlet character $\mathbb{Z} \to \mathbb{C}$.

Definition 4. Let $k(X) = c_0X^d + \cdots + c_d \in \mathbb{F}_p[X]$ be a polynomial with $c_0 \neq 0 \neq c_d$. Then $k(X)$ decomposes into linear factors over $\overline{\mathbb{F}_p}$ as $k(X) = c_0(X - t_1)\dots(X - t_d)$, and we put

$$\begin{aligned}\eta(k) :=& e_p(a(t_1 + \cdots + t_d))e_p(b(t_1^{-1} + \cdots + t_d^{-1}))\\ =& e_p(-a(c_1/c_0))e_p(-b(c_{d-1}/c_d)).\end{aligned}$$

For all other polynomials $k(X) \in \mathbb{F}_p[X]$ we put $\eta(k) := 0$.

Lemma 6. *We have*

- $|\eta(k)| \leqslant 1$ *for any* $k \in \mathbb{F}_p[X]$*;*
- $\eta(k_1k_2) = \eta(k_1)\eta(k_2)$ *for any* $k_1, k_2 \in \mathbb{F}_p[X]$.

Proof. Both statements are clear from the definition. □

Definition 5. For any integer m coprime with p, we introduce the Dirichlet series

$$L(s, \eta^m) := \sum_{k \text{ monic}} \eta^m(k)p^{-\deg(k)s}, \qquad \Re s > 1,$$

where the sum runs through the monic polynomials in $\mathbb{F}_p[X]$.

Lemma 7. *The Dirichlet series $L(s, \eta^m)$ converges absolutely and locally uniformly in the half-plane $\Re s > 1$. In addition, we have the Euler product decomposition*

$$L(s, \eta^m) = \prod_{k \text{ irred. monic}} \left(1 - \eta^m(k)p^{-\deg(k)s}\right)^{-1}, \qquad \Re s > 1,$$

which converges absolutely and locally uniformly in the half-plane $\Re s > 1$.

Proof. Let $\sigma > 1$ be fixed. In the half-plane $\Re s \geqslant \sigma$ we have, by Lemma 6,

$$\sum_{k \text{ monic}} \left|\eta^m(k)p^{-\deg(k)s}\right| \leqslant \sum_{k \text{ monic}} p^{-\deg(k)\sigma} = \sum_{d=1}^{\infty} p^{-d\sigma} \sum_{\substack{k \text{ monic} \\ \deg(k)=d}} 1 = \sum_{d=1}^{\infty} p^{d(1-\sigma)} < \infty,$$

which implies the first claim. The second claim follows from the same bound coupled with the facts that $\mathbb{F}_p[X]$ is a unique factorization domain and $\eta^m : \mathbb{F}_p[X] \to \mathbb{C}$ is completely multiplicative (cf. Lemma 6). The argument is very similar to the case of Dirichlet L-functions, hence we omit the details. $\square$

Theorem 5. *The Dirichlet series $L(s,\eta^m)$ extends to an entire function satisfying*

$$L(s,\eta^m) = 1 + S(ma,mb;p)p^{-s} + p^{1-2s}, \qquad s \in \mathbb{C}.$$

Proof. It suffices to prove that the above identity holds for $\Re s > 1$. So for the rest of the proof we assume that $\Re s > 1$, which will also take care of all convergence issues. Clearly,

$$L(s,\eta^m) = \sum_{d=1}^{\infty} p^{-ds} \sum_{\substack{k \text{ monic} \\ \deg(k)=d}} \eta^m(k),$$

hence we are led to evaluate the inner sum (cf. Definition 4). Denoting this sum by a_d, it is obvious that $a_0 = 1$, while

$$a_1 = \sum_{t \in \mathbb{F}_p} \eta^m(x-t) = \sum_{t \in \mathbb{F}_p^\times} e_p(mat)e_p(mbt^{-1}) = S(ma,mb;p).$$

Regarding a_2, we have

$$\begin{aligned} a_2 &= \sum_{c_1,c_2 \in \mathbb{F}_p} \eta^m(x^2+c_1x+c_2) = \sum_{c_1 \in \mathbb{F}_p} \sum_{c_2 \in \mathbb{F}_p^\times} e_p(-mac_1)e_p(-mbc_1/c_2) \\ &= p - 1 + \sum_{c_1 \in \mathbb{F}_p^\times} e_p(-mac_1) \sum_{c_2 \in \mathbb{F}_p^\times} e_p(-mbc_1/c_2) = p - 1 + \left(\sum_{c \in \mathbb{F}_p^\times} e_p(c)\right)^2 = p, \end{aligned}$$

while for $d \geqslant 3$ we find

$$\begin{aligned} a_d &= \sum_{c_1,\dots,c_d \in \mathbb{F}_p} \eta^m(x^d + c_1x^{d-1} + \dots + c_d) \\ &= \sum_{c_d \in \mathbb{F}_p^\times} \sum_{c_1,\dots,c_{d-1} \in \mathbb{F}_p} e_p(-mac_1)e_p(-mbc_{d-1}/c_d) = \sum_{c_d \in \mathbb{F}_p^\times} 0 = 0. \end{aligned}$$

We conclude that

$$L(s,\eta^m) = \sum_{d=1}^{\infty} a_d p^{-ds} = 1 + S(ma,mb;p)p^{-s} + p^{1-2s}, \qquad \Re s > 1.$$

The proof is complete. $\square$

Remark 2. Theorem 5 implies (and in fact is equivalent to) the functional equation

$$p^s L(s,\eta^m) = p^{1-s} L(1-s,\eta^{-m}).$$

More generally, if ω is a Hecke character of a curve of genus g over $\mathbb{F}_q$, and $\mathfrak{f}$ denotes the conductor of ω, then by Theorems 4 and 6 in [7, Chapter VII] we have

$$N^{s/2} L(s,\omega) = \kappa N^{(1-s)/2} L(1-s,\omega^{-1}),$$

where $N := q^{2g-2+\deg(\mathfrak{f})}$, and κ is a complex number of modulus 1 depending only on ω. In our case $q = p$, $g = 0$, and $\mathfrak{f} = 2(0) + 2(\infty)$ is of degree 4, so that $N = p^2$.

Theorem 6. *For any integers $1 \leqslant m \leqslant p-1$ and $n \geqslant 1$ we have*

$$-(\alpha_m^n+\beta_m^n)=\sum_{t\in\mathbb{F}_{p^n}^\times} e_p(m\,\mathrm{Tr}_n(at+bt^{-1})), \tag{8}$$

where $\mathrm{Tr}_n:\mathbb{F}_{p^n}\to\mathbb{F}_p$ is the n-trace as in Definition 3.

Proof. The idea is to analyze the logarithmic derivative of the identity

$$(1-\alpha_m p^{-s})(1-\beta_m p^{-s})=\prod_{k \text{ irred. monic}}\left(1-\eta^m(k)p^{-\deg(k)s}\right)^{-1}, \qquad \Re s>1, \tag{9}$$

that follows from Lemma 7, Theorem 5, and (2) with $T:=p^{-s}$. First of all, the left hand side is nonzero for $\Re s>1$ by the absolute convergence of the Euler product, hence $|\alpha_m|,|\beta_m|<p$ (this can also be verified directly). Therefore, on either side of (9), the factors remain in the half-plane $\Re z>0$, so that applying the principal branch of the logarithm on this half-plane yields

$$\log(1-\alpha_m p^{-s})+\log(1-\beta_m p^{-s})=\sum_{k \text{ irred. monic}}-\log\left(1-\eta^m(k)p^{-\deg(k)s}\right), \qquad \Re s>1.$$

Indeed, the two sides can only differ by a (constant) multiple of $2\pi i$, and then letting $s>1$ and $s\to\infty$ shows that the difference is zero. We expand the logarithmic values via

$$\log(1-z)=-\sum_{n=1}^\infty\frac{z^n}{n}, \qquad |z|<1,$$

and arrive at

$$\sum_{n=1}^\infty\frac{-(\alpha_m^n+\beta_m^n)p^{-ns}}{n}=\sum_{k \text{ irred. monic}}\sum_{r=1}^\infty\frac{\eta^{mr}(k)p^{-r\deg(k)s}}{r}, \qquad \Re s>1.$$

Both sides converge absolutely and locally uniformly, hence we can differentiate termwise and divide by $-\log p$ to obtain

$$\sum_{n=1}^\infty-(\alpha_m^n+\beta_m^n)p^{-ns}=\sum_{k \text{ irred. monic}}\sum_{r=1}^\infty\deg(k)\eta^{mr}(k)p^{-r\deg(k)s}, \qquad \Re s>1.$$

By comparing the Dirichlet coefficients on the two sides, we infer that

$$-(\alpha_m^n+\beta_m^n)=\sum_{\substack{k \text{ irred. monic}\\ r\deg(k)=n}}\deg(k)\eta^{mr}(k), \qquad n\geqslant 1.$$

In other words,

$$-(\alpha_m^n+\beta_m^n)=\sum_{d\mid n}\sum_{\substack{k \text{ irred. monic}\\ \deg(k)=d}}d\eta^{\frac{mn}{d}}(k), \qquad n\geqslant 1. \tag{10}$$

The polynomial $k(X)=X$ does not contribute to the inner sum, while the other irreducible monic polynomials $k\in\mathbb{F}_p[X]$ correspond bijectively to the Frobenius orbits lying in $\mathbb{F}_{p^n}^\times$ (cf. Theorem 3 and the remarks preceding it). Namely, if $\{t_1,\dots,t_d\}$ is the set of roots of k in $\overline{\mathbb{F}_p}$, then $\{t_1,\dots,t_d\}\subset\mathbb{F}_{p^n}^\times$ is the corresponding Frobenius orbit of size $d\mid n$, and we have (cf. Definition 4)

$$\eta^{\frac{mn}{d}}(k)=e_p\left(ma\frac{n}{d}(t_1+\dots+t_d)\right)e_p\left(mb\frac{n}{d}(t_1^{-1}+\dots+t_d^{-1})\right).$$

For any $1\leqslant j\leqslant d$, we can interpret (cf. Definition 3)

$$\frac{n}{d}(t_1+\dots+t_d)=\frac{n}{d}\sum_{i=0}^{d-1}\sigma^i(t_j)=\sum_{i=0}^{n-1}\sigma^i(t_j)=\mathrm{Tr}_n(t_j)$$

and

$$\frac{n}{d}(t_1^{-1}+\dots+t_d^{-1})=\frac{n}{d}\sum_{i=0}^{d-1}\sigma^i(t_j^{-1})=\sum_{i=0}^{n-1}\sigma^i(t_j^{-1})=\mathrm{Tr}_n(t_j^{-1}),$$

hence

$$\eta^{\frac{mn}{d}}(k) = e_p(ma\,\mathrm{Tr}_n(t_j))e_p(mb\,\mathrm{Tr}_n(t_j^{-1})) = e_p(m\,\mathrm{Tr}_n(at_j + bt_j^{-1})), \qquad 1 \leqslant j \leqslant d.$$

Summing up these equations for $1 \leqslant j \leqslant d$, we get

$$d\eta^{\frac{mn}{d}}(k) = \sum_{j=1}^{d} e_p(m\,\mathrm{Tr}_n(at_j + bt_j^{-1})).$$

The right hand side is the sum of $e_p(m\,\mathrm{Tr}_n(at + bt^{-1}))$ over the Frobenius orbit $\{t_1, \ldots, t_d\}$ corresponding to k, hence (10) readily implies (8). □

Corollary 3. *The identity* (3) *holds for any positive integer n.*

Proof. Using Theorems 6 and 4, we calculate

$$\begin{aligned}
p^n - 1 - \sum_{m=1}^{p-1}(\alpha_m^n + \beta_m^n) &= \sum_{m=0}^{p-1} \sum_{t \in \mathbb{F}_{p^n}^\times} e_p(m\,\mathrm{Tr}_n(at + bt^{-1})) \\
&= \sum_{t \in \mathbb{F}_{p^n}^\times} \sum_{m=0}^{p-1} e_p(m\,\mathrm{Tr}_n(at + bt^{-1})) \\
&= p|\{t \in \mathbb{F}_{p^n}^\times : \ \mathrm{Tr}_n(at + bt^{-1}) = 0\}| \\
&= |\{(x,t) \in \mathbb{F}_{p^n} \times \mathbb{F}_{p^n}^\times : \ x^p - x = at + bt^{-1}\}| \\
&= |\{(x,t) \in \mathbb{F}_{p^n} \times \mathbb{F}_{p^n} : \ at^2 - t(x^p - x) + b = 0\}| \\
&= |\{(x,t) \in \mathbb{F}_{p^n} \times \mathbb{F}_{p^n} : \ (2at - (x^p - x))^2 = (x^p - x)^2 - 4ab\}| \\
&= |\{(x,y) \in \mathbb{F}_{p^n} \times \mathbb{F}_{p^n} : \ y^2 = (x^p - x)^2 - 4ab\}|.
\end{aligned}$$

Comparing the two sides, we obtain (3). □

Remark 3. The equation $y^2 = (x^p - x)^2 - 4ab$ defines an affine real hyperelliptic curve of genus $p-1$ over $\mathbb{F}_p$. It has two points at infinity, so by (3) the number of $\mathbb{F}_{p^n}$-rational points of the completed (nonsingular projective) curve C equals

$$|C(\mathbb{F}_{p^n})| = p^n + 1 - \sum_{m=1}^{p-1}(\alpha_m^n + \beta_m^n).$$

An elegant way of expressing this fact is that the zeta function of C equals

$$\zeta_C(s) = \frac{\prod_{m=1}^{p-1}(1 - \alpha_m p^{-s})(1 - \beta_m p^{-s})}{(1 - p^{-s})(1 - p^{1-s})} = \zeta_P(s) \prod_{m=1}^{p-1} L(s, \eta^m),$$

where P is the projective line over $\mathbb{F}_p$.

4. The Hasse derivative

In the light of Corollary 3, we have reduced Theorem 1 to the statement that the equation $y^2 = (x^p - x)^2 - 4ab$ has $p^n + O_p(p^{n/2})$ solutions over the finite field $\mathbb{F}_{p^n}$. Recall that $p > 2$ is a fixed odd prime, and ab is coprime to p. More generally, we shall prove using the method of Stepanov [4] the following bound for hyperelliptic curves over finite fields, itself a special case of Weil's theorem for all algebraic curves over finite fields [6, p. 70].

Theorem 7 (Weil, Stepanov)**.** *Let $q = p^n$ be an odd prime power, and let $f(X) \in \mathbb{F}_q[X]$ be a polynomial of degree $m \geqslant 3$. Assume that $q > 6m$ and $f(X)$ is not a complete square in $\overline{\mathbb{F}_p}[X]$. If N denotes the number of solutions of the equation $y^2 = f(x)$ over $\mathbb{F}_q$, then*

$$|N - q| < 4m\lceil\sqrt{q}\,\rceil. \tag{11}$$

Remark 4. Using the functional equation for the L-function associated with the hyperelliptic curve $y^2 = f(x)$ over $\mathbb{F}_q$ and its extensions $\mathbb{F}_{q^\nu}$, one can deduce that the above bound improves itself to

$$|N-q| \leqslant 2\left\lfloor \frac{m-1}{2} \right\rfloor \sqrt{q} < m\sqrt{q},$$

even without the assumption $q > 6m$. See Lemma 4 in [7, Appendix V] for more detail.

By Lemma 4, the multiplicative group $\mathbb{F}_q^\times$ is cyclic of even order $q-1$, hence for any $t \in \mathbb{F}_q^\times$ we have $t^{\frac{q-1}{2}} = 1$ or $t^{\frac{q-1}{2}} = -1$ depending on whether t is a square in $\mathbb{F}_q^\times$ or not. Moreover, every square in $\mathbb{F}_q^\times$ is a square in precisely two ways, hence with the notation

$$N_a := |\{x \in \mathbb{F}_q : \ f(x)^{\frac{q-1}{2}} = a\}|, \qquad a \in \{0, \pm 1\},$$

we can express the defect $N-q$ as

$$N - q = (N_0 + 2N_1) - (N_0 + N_1 + N_{-1}) = N_1 - N_{-1}. \tag{12}$$

In other words, Theorem 7 bounds the difference between the number of $x \in \mathbb{F}_q$ with $f(x)$ a nonzero square and those with $f(x)$ not a square.

Now the proof of Theorem 7 relies on two basic ideas. The first idea is that it suffices to show the one-sided bound

$$\max(N_0 + N_1, N_0 + N_{-1}) < \frac{q}{2} + 2m\lceil \sqrt{q}\,\rceil. \tag{13}$$

Indeed, this inequality readily yields

$$\max(N_1, N_{-1}) < \frac{q}{2} + 2m\lceil \sqrt{q}\,\rceil,$$

and by $N_0 + N_1 + N_{-1} = q$ also

$$\min(N_1, N_{-1}) = q - \max(N_0 + N_{-1}, N_0 + N_1) > \frac{q}{2} - 2m\lceil \sqrt{q}\,\rceil,$$

whence (11) follows via (12):

$$|N-q| = |N_1 - N_{-1}| = \max(N_1, N_{-1}) - \min(N_1, N_{-1}) < 4m\lceil \sqrt{q}\,\rceil.$$

The second idea is to exhibit, for any $a \in \{\pm 1\}$ and a suitable integer $\ell \geqslant 1$, a nonzero polynomial $h_a(X) \in \mathbb{F}_q[X]$ such that any $x \in \mathbb{F}_q$ satisfying $f(x)^{\frac{q-1}{2}} \in \{0, a\}$ is a root of $h_a(X)$ of order at least ℓ, i.e. $(X-x)^\ell$ divides $h_a(X)$ in $\mathbb{F}_q[X]$. The point is that in this case we have

$$\ell(N_0 + N_a) \leqslant \deg h_a, \qquad a \in \{\pm 1\}, \tag{14}$$

and by optimizing ℓ in terms of q and m we can deduce (13), hence also Theorem 7.

In order to verify the divisibility relation $(X-x)^\ell \mid h_a(X)$ in $\mathbb{F}_q[X]$, we introduce a simple but powerful tool, the *Hasse derivative*.

Definition 6. Let F be a field, and let $h(X) \in F[X]$ be any polynomial. In the ring of polynomials of two variables $F[X,Y]$, there is a unique decomposition

$$h(X+Y) = \sum_{k=0}^{\infty} (E^k h)(X) Y^k, \tag{15}$$

where $(E^k h)(X) \in F[X]$, and the terms for $k > \deg h$ vanish. The polynomial $(E^k h)(X)$ is called the k-th Hasse derivative of $h(X)$.

It is clear that the operator $E^k : F[X] \to F[X]$ is F-linear, and also translation invariant in the sense that for any $x \in F$ the k-th Hasse derivative of the translated polynomial $h(X+x)$ equals $(E^k h)(X+x)$. It is also clear that $\deg(E^k h) \leqslant (\deg h) - k$ for $0 \leqslant k \leqslant \deg h$, with the convention that $\deg 0 = 0$, while $E^k h = 0$ for $k > \deg h$. In fact the binomial theorem gives that $E^k(X^n) = \binom{n}{k} X^{n-k}$ for $0 \leqslant k \leqslant n$, while $E^k(X^n) = 0$ for $k > n$.

Lemma 8. *Let F be a field, $h(X) \in F[X]$, and $x \in F$. Then $(X-x)^\ell \mid h(X)$ holds in $F[X]$ if and only if $(E^k h)(x) = 0$ for any $0 \leqslant k < \ell$.*

Proof. By translation invariance, we can assume without loss of generality that $x = 0$. Then, (15) implies by the substitution $X \mapsto 0$ that

$$h(Y) = \sum_{k=0}^{\infty} (E^k h)(0) Y^k,$$

whence $Y^\ell \mid h(Y)$ holds in $F[Y]$ if and only if $(E^k h)(0) = 0$ for any $0 \leqslant k < \ell$. □

Lemma 9 (Leibniz rule)**.** *For any polynomials $h_1(X), \dots, h_n(X) \in F[X]$ we have*

$$E^k(h_1 \cdots h_n) = \sum_{\substack{k_1 + \cdots + k_n = k \\ k_1, \dots, k_n \geqslant 0}} E^{k_1}(h_1) \dots E^{k_n}(h_n).$$

Proof. This is straightforward from the definition (15). Indeed,

$$\begin{aligned} h_1(X+Y) \dots h_n(X+Y) &= \left(\sum_{k_1=0}^{\infty} (E^{k_1} h_1)(X) Y^{k_1} \right) \cdots \left(\sum_{k_n=0}^{\infty} (E^{k_n} h_n)(X) Y^{k_n} \right) \\ &= \sum_{k_1, \dots, k_n \geqslant 0} \Big(E^{k_1}(h_1)(X) \dots E^{k_n}(h_n)(X) \Big) Y^{k_1 + \cdots + k_n}, \end{aligned}$$

and the result follows. □

Lemma 10. *Let F be a field, and let $f(X), g(X) \in F[X]$ be arbitrary. For any integers $0 \leqslant k < n$, the polynomial $E^k(gf^n)$ is of the form $g^{(k)} f^{n-k}$, where $g^{(k)}(X) \in F[X]$. Moreover, for a fixed f, the polynomial $g^{(k)}$ depends F-linearly on g. Finally,*

$$\deg g^{(k)} \leqslant \deg g + k \deg f - k. \tag{16}$$

Proof. By Lemma 9,

$$E^k(gf^n) = \sum_{\substack{k_0 + k_1 + \cdots + k_n = k \\ k_0, k_1, \dots, k_n \geqslant 0}} E^{k_0}(g) E^{k_1}(f) \dots E^{k_n}(f).$$

Clearly, at least $n-k$ of the integers $k_1, \dots, k_n \geqslant 0$ must vanish, hence each term on the right hand side is divisible by f^{n-k} in $F[X]$. This shows that $E^k(gf^n)$ is of the form $g^{(k)} f^{n-k}$, where $g^{(k)}(X) \in F[X]$. Moreover, for a fixed f, the factor $E^{k_0}(g)$ depends F-linearly on g, hence the same is true of the polynomial $g^{(k)}$. Finally, (16) is immediate from

$$\deg(g^{(k)} f^{n-k}) \leqslant \deg(gf^n) - k.$$

□

5. Stepanov's auxiliary polynomials

In this section we construct the two nonzero auxiliary polynomials $h_{\pm 1}(X) \in \mathbb{F}_q[X]$ that will allow us to derive (13) via (14). We assume the conditions of Theorem 7, and we fix a value $a \in \{\pm 1\}$. The statement of Theorem 7 does not change upon replacing $f(X)$ by $f(X+x)$ for any $x \in \mathbb{F}_q$, hence we can assume without loss of generality that $f(0) \neq 0$. Indeed, $f(x) \neq 0$ for some $x \in \mathbb{F}_q$, because $f(X)$ has degree less than q.

We have seen that for the validity of (14) it suffices that

$$f(x)^{\frac{q-1}{2}} \in \{0, a\} \implies (E^k h_a)(x) = 0, \qquad x \in \mathbb{F}_q, \ 0 \leqslant k < \ell. \tag{17}$$

We choose $h_a(X)$ to be a multiple of $f(X)^\ell$, so that we can restrict to the values $f(x)^{\frac{q-1}{2}} = a$ in (17). Specifically, we seek $h_a(X)$ in the form

$$h_a(X) := f(X)^\ell \sum_{0 \leqslant j < J} \left\{ r_j(X) + s_j(X) f(X)^{\frac{q-1}{2}} \right\} X^{jq}, \tag{18}$$

where $J>0$ is a real parameter (to be chosen later in terms of ℓ, m, q), and

$$r_j(X), s_j(X) \in \mathbb{F}_q[X], \qquad 0 \leqslant j < J$$

are any polynomials with

$$\deg r_j, \deg s_j < \frac{q-m}{2}, \qquad 0 \leqslant j < J. \tag{19}$$

We examine first the possibility that $h_a(X)=0$. Assume that this is the case, but not all the polynomials $r_j(X), s_j(X) \in \mathbb{F}_q[X]$ are zero. Let $0 \leqslant i < J$ be minimal such that either $r_i(X)$ or $s_i(X)$ is nonzero. Then

$$\sum_{i \leqslant j < J} \left\{ r_j(X) + s_j(X) f(X)^{\frac{q-1}{2}} \right\} X^{(j-i)q} = 0,$$

whence in $\mathbb{F}_q[X]$ we have the congruence

$$r_i(X) + s_i(X) f(X)^{\frac{q-1}{2}} \equiv 0 \pmod{X^q}.$$

From here we infer that

$$r_i(X)^2 f(X) \equiv s_i(X)^2 f(X)^q \equiv s_i(X)^2 f(X^q) \equiv s_i(X)^2 f(0) \pmod{X^q}.$$

By (19), the two sides are polynomials of degree less than q, hence in fact

$$r_i(X)^2 f(X) = s_i(X)^2 f(0).$$

As $f(0) \neq 0$, both $r_i(X)$ and $s_i(X)$ are nonzero, and $f(X)$ is a complete square in $\overline{\mathbb{F}_p}[X]$. This contradicts the assumptions of Theorem 7, hence we proved that $h_a(X) \neq 0$ unless all the polynomials $r_j(X), s_j(X) \in \mathbb{F}_q[X]$ are zero.

Now we examine what $(E^k h_a)(x) = 0$ means for $f(x)^{\frac{q-1}{2}} = a$ and $0 \leqslant k < \ell$, cf. (17). In order to find the Hasse derivative $(E^k h_a)(X)$, we go back to the definition (15), and we make a simple observation. Starting from the congruence in $\mathbb{F}_q[X,Y]$,

$$(X+Y)^{jq} = (X^q + Y^q)^j \equiv X^{jq} \pmod{Y^q},$$

we see that

$$h_a(X+Y) \equiv \sum_{0 \leqslant j < J} \left\{ r_j(X+Y) f(X+Y)^{\ell} + s_j(X+Y) f(X+Y)^{\ell + \frac{q-1}{2}} \right\} X^{jq} \pmod{Y^q},$$

whence for $0 \leqslant k < q$ the coefficient of Y^k as an element of $\mathbb{F}_q[X]$ must be the same on the two sides. That is,

$$(E^k h_a)(X) = \sum_{0 \leqslant j < J} \left\{ E^k(r_j f^{\ell})(X) + E^k(s_j f^{\ell + \frac{q-1}{2}})(X) \right\} X^{jq}, \qquad 0 \leqslant k < q.$$

By Lemma 10, we can rewrite this identity as

$$(E^k h_a)(X) = f(X)^{\ell - k} \sum_{0 \leqslant j < J} \left\{ r_j^{(k)}(X) + s_j^{(k)}(X) f(X)^{\frac{q-1}{2}} \right\} X^{jq}, \qquad 0 \leqslant k < q, \tag{20}$$

where the polynomials $r_j^{(k)}, s_j^{(k)} \in \mathbb{F}_q[X]$ depend $\mathbb{F}_q$-linearly on the initial $r_j, s_j \in \mathbb{F}_q[X]$, and

$$\deg r_j^{(k)}, \deg s_j^{(k)} < \frac{q-m}{2} + k(m-1), \qquad 0 \leqslant j < J, \ \ 0 \leqslant k < q. \tag{21}$$

In passing, it is worthwhile to remark that $r_j^{(0)} = r_j$ and $s_j^{(0)} = s_j$. From now on we assume that $\ell \leqslant q$, then by (20) we can reduce (17) to the simpler condition

$$\sum_{0 \leqslant j < J} \left\{ r_j^{(k)}(X) + a s_j^{(k)}(X) \right\} X^j = 0, \qquad 0 \leqslant k < \ell. \tag{22}$$

Here we relied on the crucial fact that $x^{jq} = x^j$ for any $x \in \mathbb{F}_q$.

The constraints (22) constitute a homogeneous system of linear equations for the coefficients of $r_j(X)$ and $s_j(X)$. By (19), the number of variables in this system is

$$\geqslant 2J\left\lceil\frac{q-m}{2}\right\rceil \geqslant J(q-m),$$

while by (21), the number of equations is

$$\leqslant \sum_{0\leqslant k<\ell}\left\lceil\frac{q-m}{2}+k(m-1)+J\right\rceil < \ell\left(\frac{q-m}{2}+J\right)+\frac{\ell^2}{2}(m-1).$$

This means that the construction (18) yields a nonzero polynomial $h_a(X)\in\mathbb{F}_q[X]$ validating (14) as long as $\ell\leqslant q$ and

$$J(q-m)\geqslant \ell\left(\frac{q-m}{2}+J\right)+\frac{\ell^2}{2}(m-1).$$

Rearranging the last inequality,

$$\left(J-\frac{\ell}{2}\right)(q-m-\ell)\geqslant\frac{\ell^2 m}{2},$$

hence by imposing $\ell\leqslant q/3$ and utilizing $m<q/6$ (cf. Theorem 7) it suffices to have

$$\left(J-\frac{\ell}{2}\right)\frac{q}{2}\geqslant\frac{\ell^2 m}{2}.$$

This motivates the choice

$$J:=\frac{\ell}{2}+\frac{\ell^2 m}{q}.$$

With this choice (14) yields, upon recalling (18) and (19),

$$\ell(N_0+N_a)\leqslant\deg h_a<m\left(\ell+\frac{q-1}{2}\right)+\frac{q-m}{2}+Jq<mq+\frac{\ell q}{2}+\ell^2 m.$$

In short,

$$N_0+N_a<\frac{q}{2}+\frac{mq}{\ell}+\ell m,$$

and by choosing $\ell:=\lceil\sqrt{q}\rceil$ we obtain (13). Note that the intermediate constraint $\ell\leqslant q/3$ is now automatically satisfied, because the conditions of Theorem 7 force $q>18$.

The proof of Theorem 7 is now complete. To conclude Theorem 1 via Corollary 3, we apply Theorem 7 for $n\geqslant 4$ and $f(X):=(X^p-X)^2-4ab$. All we need to check is that $f(X)$ is not a complete square in $\overline{\mathbb{F}_p}[X]$. However, this is clear: $f(X)=g(X)^2$ would imply

$$(X^p-X-g(X))(X^p-X+g(X))=4ab,$$

an obvious contradiction to the fact that one of the factors $X^p-X\pm g(X)$ is non-constant.

6. Supplements

With a bit of algebraic number theory, we can show that the inequality in Theorem 1 is always strict. The proof below is due to Elkies and MathOverflow user Lucia (see [3]).

Theorem 8. *Let $p>2$ be a prime, and let $(ab,p)=1$. Then $|S(a,b;p)|<2\sqrt{p}$.*

Proof. The Kloosterman sum $S(a,b;p)$ is real, as can be seen by writing $-t$ for t in (1). Therefore, by Theorem 1, we only need to exclude the possibility that

$$\sum_{t=1}^{p-1} e_p(at+b\bar{t})=\pm 2\sqrt{p}. \tag{23}$$

Let us assume (23). Then both sides lie in the ring $\mathbb{Z}[\xi]$, where $\xi := e^{2\pi i/p}$, which consists of the integral linear combinations of $1, \xi, \xi^2, \dots$. Raising the equation to the p-th power yields, by the multinomial theorem,

$$\sum_{t=1}^{p-1} 1 \equiv \pm 2^p p^{p/2} \pmod{p\mathbb{Z}[\xi]}.$$

The left hand side is congruent to -1 modulo $p\mathbb{Z}[\xi]$, hence further squaring both sides,

$$1 \equiv 2^{2p} p^p \equiv 0 \pmod{p\mathbb{Z}[\xi]}.$$

That is, $1 \in p\mathbb{Z}[\xi]$, which is a contradiction as we explain now. It is classical and easy to prove with the Schönemann–Eisenstein criterion that the cyclotomic polynomial

$$k(X) := X^{p-1} + X^{p-2} + \dots + X + 1 = (X - \xi)(X - \xi^2) \dots (X - \xi^{p-1})$$

is irreducible over $\mathbb{Q}$, hence $\mathbb{Q}(\xi)$ is isomorphic to $\mathbb{Q}[X]/(k(X))$ by Lemma 2 and its proof. In particular, $\{1, \xi, \dots, \xi^{p-2}\}$ is a basis of $\mathbb{Q}(\xi)$ as a vector space over $\mathbb{Q}$, and $\mathbb{Z}[\xi]$ consists of the vectors whose coordinates are integers with respect to this basis. This shows readily that $1 \notin p\mathbb{Z}[\xi]$, because $1 \notin p\mathbb{Z}$, and we are done. □

Finally, following Heath-Brown [1], we give an application of Theorem 1 to the distribution of products modulo a prime number.

Theorem 9. *Let $p > 2$ be a prime number. Let $\mathscr{U}, \mathscr{V} \subseteq \{1, 2, \dots, p-1\}$ be two intervals, and let $r \in \{1, 2, \dots, p-1\}$ be a nonzero residue modulo p. Then*

$$\left| \sum_{\substack{u \in \mathscr{U},\ v \in \mathscr{V} \\ uv \equiv r \pmod{p}}} 1 - \frac{|\mathscr{U}||\mathscr{V}|}{p-1} \right| < 2p^{1/2} (\log p)^2.$$

Proof. Using Fourier analysis on $\mathbb{Z}/p\mathbb{Z}$, we can express

$$\begin{aligned}
\sum_{\substack{u \in \mathscr{U},\ v \in \mathscr{V} \\ uv \equiv r \pmod{p}}} 1 &= \sum_{t=1}^{p-1} \Bigg(\sum_{\substack{u \in \mathscr{U} \\ t \equiv u \pmod{p}}} 1 \Bigg) \Bigg(\sum_{\substack{v \in \mathscr{V} \\ \bar{t} \equiv \bar{r} v \pmod{p}}} 1 \Bigg) \\
&= \sum_{t=1}^{p-1} \left(\sum_{u \in \mathscr{U}} \frac{1}{p} \sum_{a=1}^{p} e_p(a(t-u)) \right) \left(\sum_{v \in \mathscr{V}} \frac{1}{p} \sum_{b=1}^{p} e_p(b(\bar{t} - \bar{r} v)) \right) \\
&= \frac{1}{p^2} \sum_{a,b=1}^{p} \left(\sum_{t=1}^{p-1} e_p(at + b\bar{t}) \right) \left(\sum_{u \in \mathscr{U}} e_p(-au) \right) \left(\sum_{v \in \mathscr{V}} e_p(-b\bar{r} v) \right).
\end{aligned}$$

The first inner sum is $p-1$ when both a and b equal p, it is -1 when exactly one of a and b equals p, and otherwise it is the Kloosterman sum $S(a, b; p)$ considered in Theorem 1. Using this information, we obtain

$$\sum_{\substack{u \in \mathscr{U},\ v \in \mathscr{V} \\ uv \equiv r \pmod{p}}} 1 = |\mathscr{U}||\mathscr{V}| \frac{p+1}{p^2} + \frac{1}{p^2} \sum_{a,b=1}^{p-1} S(a, b; p) \left(\sum_{u \in \mathscr{U}} e_p(-au) \right) \left(\sum_{v \in \mathscr{V}} e_p(-b\bar{r} v) \right),$$

whence by Theorem 1 and the fact that $\mathscr{U}$ and $\mathscr{V}$ are intervals,

$$\begin{aligned}
\left| \sum_{\substack{u \in \mathscr{U},\ v \in \mathscr{V} \\ uv \equiv r \pmod{p}}} 1 - \frac{|\mathscr{U}||\mathscr{V}|}{p-1} \right| &< \frac{1}{p} + 2\sqrt{p} \left(\frac{1}{p} \sum_{a=1}^{p-1} \left| \sum_{u \in \mathscr{U}} e_p(-au) \right| \right) \left(\frac{1}{p} \sum_{b=1}^{p-1} \left| \sum_{v \in \mathscr{V}} e_p(-b\bar{r} v) \right| \right) \\
&< \frac{1}{p} + 2p^{1/2} \left(\frac{1}{p} \sum_{c=1}^{p-1} \frac{1}{\sin\left(\frac{\pi c}{p}\right)} \right)^2 < 2p^{1/2} (\log p)^2.
\end{aligned}$$

Here, the last inequality can be checked numerically for $p<11$, while for $p\geqslant 11$ we verify it as follows. We have

$$\frac{1}{p}\sum_{c=1}^{p-1}\frac{1}{\sin\left(\frac{\pi c}{p}\right)}=\frac{2}{p}\sum_{c=1}^{\frac{p-1}{2}}\frac{1}{\sin\left(\frac{\pi c}{p}\right)}<\sum_{c=1}^{\frac{p-1}{2}}\frac{1}{c}<0.68+\log\frac{p-1}{2}<-0.01+\log p,$$

therefore

$$\left(\frac{1}{p}\sum_{c=1}^{p-1}\frac{1}{\sin\left(\frac{\pi c}{p}\right)}\right)^2<(-0.01+\log p)^2<(\log p)^2-\frac{\log p}{100}<(\log p)^2-\frac{1}{2p^{3/2}}.$$

The proof is complete. $\square$

Corollary 4. *Let $p,r,\mathscr{U},\mathscr{V}$ as in Theorem 9. If $|\mathscr{U}||\mathscr{V}|>2p^{3/2}(\log p)^2$, then the congruence $uv\equiv r \pmod p$ has a solution in $u\in\mathscr{U}$ and $v\in\mathscr{V}$.*

Proof. If the congruence $uv\equiv r \pmod p$ has no solution in $u\in\mathscr{U}$ and $v\in\mathscr{V}$, then Theorem 9 yields

$$\frac{|\mathscr{U}||\mathscr{V}|}{p-1}<2p^{1/2}(\log p)^2,$$

hence also $|\mathscr{U}||\mathscr{V}|<2p^{3/2}(\log p)^2$. $\square$

References

[1] D. R. Heath-Brown, *Arithmetic applications of Kloosterman sums*, Nieuw Arch. Wiskd. (5) **1** (2000), 380–384.

[2] H. Iwaniec, E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, 53, American Mathematical Society, Providence, RI, 2004.

[3] Responses to MathOverflow question No. 166297, http://mathoverflow.net/questions/166297

[4] S. A. Stepanov, *An elementary proof of the Hasse-Weil theorem for hyperelliptic curves*, J. Number Theory **4** (1972), 118–143.

[5] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci. U. S. A. **34** (1948), 204–207.

[6] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Actualités Sci. Ind., 1041, Hermann et Cie., Paris, 1948.

[7] A. Weil, *Basic number theory*, Third edition, Die Grundlehren der Mathematischen Wissenschaften, 144, Springer-Verlag, New York-Berlin, 1974.

MTA Alfréd Rényi Institute of Mathematics, POB 127, Budapest H-1364, Hungary
E-mail address: `gharcos@renyi.hu`

Central European University, Nador u. 9, Budapest H-1051, Hungary
E-mail address: `harcosg@ceu.hu`