

Weil Sums over Small Subgroups

BY ALINA OSTAFE AND IGOR E. SHPARLINSKI

School of Mathematics and Statistics, University of New South Wales, Sydney NSW 2052, Australia.

e-mails: alina.ostafe@unsw.edu.au, igor.shparlinski@unsw.edu.au

JOSÉ FELIPE VOLOCH

School of Mathematics and Statistics, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand.

e-mail: felipe.voloch@canterbury.ac.nz

(Received 23 November 2022; revised 05 July 2023; accepted 03 July 2023)

Abstract

We obtain new bounds on short Weil sums over small multiplicative subgroups of prime finite fields which remain nontrivial in the range the classical Weil bound is already trivial. The method we use is a blend of techniques coming from algebraic geometry and additive combinatorics.

2020 Mathematics Subject Classification: 11T23 (Primary); 11D79, 11L07 (Secondary)

1. Introduction

1.1. Set-up and motivation

Let p be a prime. Given a subset \mathcal{X} of a finite field \mathbb{F}_p of p elements and a polynomial $f \in \mathbb{F}_p[X]$, we define the *Weil sum* over \mathcal{X} as

$$S(\mathcal{X}; f) = \sum_{x \in \mathcal{X}} \mathbf{e}_p(f(x)),$$

where

$$\mathbf{e}_p(z) = \exp(2\pi iz/p),$$

and we always assume that the elements of \mathbb{F}_p are represented by the set $\{0, \dots, p-1\}$.

The celebrated result of Weil [28] asserts that for any nontrivial polynomial $f \in \mathbb{F}_p[X]$, when $\mathcal{X} = \mathbb{F}_p$, we have

$$|S(\mathbb{F}_p; f)| \leq (n-1)p^{1/2}, \tag{1.1}$$

where $n = \deg f$, see also, for example, [8, chapter 11] and [14, chapter 6].

The sums $S(\mathbb{F}_p; f)$ are usually called *complete sums*. The problem usually becomes harder for smaller sets \mathcal{X} , that is, for sums called *incomplete sums*.

Most of the attention the incomplete sums received is in the case of the sets $\mathcal{I}_N = \{0, \dots, N - 1\}$ of $N \leq p$ consecutive integers. In fact, using the classical Weil bound and the completing technique (see [8, section 12.2]), it is easy to show that in this case

$$S(\mathcal{I}_N; f) = O(p^{1/2} \log p) \tag{1.2}$$

for any nonlinear polynomial $f \in \mathbb{F}_p[X]$, where the implied constant depends only on the degree n . Clearly the bound (1.2) is nontrivial only for $N \geq p^{1/2+\varepsilon}$ for some fixed $\varepsilon > 0$. For smaller values of N one can also use general bounds on the Weyl sums, see, for example, [3, theorem 5], which remain nontrivial as long as $N \geq p^{1/n+\varepsilon}$, which is an optimal range. In the special case of monomials $f(X) = X^n$, that is, for *Gauss sums*, Kerr [9] has obtained a better bound in the middle range of N . Kerr and Macourt [10, theorem 1.4] have also considered exponential sums over *generalised arithmetic progressions* rather than over intervals.

The multiplicative analogues of this problem, when, instead of interval, the sum is over a multiplicative subgroup $\mathcal{G} \subseteq \mathbb{F}_p^*$ has also been studied, however significantly less general results are known. Again, the Weil bound (1.1), using that

$$S(\mathcal{G}; f) = \frac{\tau}{p-1} \sum_{x=1}^{p-1} \mathbf{e}_p \left(f \left(x^{(p-1)/\tau} \right) \right) = O(p^{1/2}), \tag{1.3}$$

instantly gives a nontrivial result for subgroups of order $\tau = \#\mathcal{G} \geq p^{1/2+\varepsilon}$.

In the case of linear polynomials, the bound of [23, theorem 2] has started a series of further improvements which goes beyond this limitation on $\#\mathcal{G}$, see [2, 4, 5, 7, 11, 21] and references therein.

Significantly less is known in the case of non-linear polynomials $f \in \mathbb{F}_p[X]$. Until very recently, the only known approach to such bounds was that of Bourgain [1], which actually works in a much more general scenario of exponential sums with linear combinations of several exponential functions. This result of Bourgain [1] gives a bound saving some power p^η compared to the trivial bound, however the exponent η is not explicit and an attempt to make it explicit in [20, theorems 4 and 5] has some problems. Indeed, it seems that the argument in [20] quotes incorrectly the result of [22, corollary 16], which, after correcting, leads to exponentially smaller saving.

More recently, the authors [19] have used a different approach to a similar problem, based on a bound for the number of rational points on curves over finite fields from [27], which in some cases is stronger than the use of the classical Weil bound [15, equation (5.7)], and estimate Kloosterman sums

$$K(\mathcal{G}; a, b) = \sum_{x \in \mathcal{G}} \mathbf{e}_p \left(ax + bx^{-1} \right)$$

over a subgroup \mathcal{G} of order τ with $a, b \in \mathbb{F}_p^*$. More precisely, the Weil bound, in the form given for example, in [17, theorem 2], instantly gives

$$K(\mathcal{G}; a, b) = \frac{\tau}{p-1} \sum_{x=1}^{p-1} \mathbf{e}_p \left(ax^{(p-1)/\tau} + bx^{-(p-1)/\tau} \right) = O \left(p^{1/2} \right)$$

for $a, b \in \mathbb{F}_p^*$, which becomes trivial for $\tau < p^{1/2}$.

On the other hand, by [19, corollary 2.9] we have

$$K(\mathcal{G}; a, b) \ll \tau^{20/27} p^{1/9}, \quad (1.4)$$

where, as usual, the notations $U = O(V)$, $U \ll V$ and $V \gg U$ are equivalent to $|U| \leq cV$ for some positive constant c , which throughout this work may depend only on n (and thus is absolute in (1.4)). Clearly, the bound (1.4) is nontrivial for $\tau \geq p^{3/7+\varepsilon}$ for some fixed $\varepsilon > 0$.

We also note that several other bounds of exponential sums which rely on the result of [27] and go beyond the Weil bound (1.1) have been given in [24], see also [25].

1.2. Approach

It is important to recall that it has been shown in [23], and used again in [7], that in the case of linear polynomials, good bounds on the 4th moment of the corresponding sums already allow us to improve the Weil bound (1.3).

However, for higher degree polynomials this is not sufficient and one needs strong bounds on at least the 6th moment. Hence, to obtain nontrivial bounds for $S(\mathcal{G}; f)$ for a given non-constant polynomial $f \in \mathbb{F}_p[X]$ and a small multiplicative subgroup \mathcal{G} of \mathbb{F}_p^* , we modify the ideas of our previous work [19] to investigate some high degree systems of polynomial equations. The main difficulty here is to study the *generic* absolute irreducibility of a certain family of curves and to be able to apply the result of [27, theorem (i)]. We then combine this with the inductive approach of Bourgain [1].

More precisely, the method of Bourgain [1] (see also the exposition in [6, section 4.4]) is inductive on the number of non-constant terms r of the polynomial, and it requires the case $r = 1$ and $r = 2$ as the basis of induction. For $r = 1$ we can use the bound of Shkredov [21], or even one of the earlier bounds from [7, 23]. So we start with obtaining a bound for binomial sums over a subgroup, see Lemma 3.7, which is similar to (1.4). The proof of Lemma 3.7 resembles that of our previous work [19, theorem 2.7] but requires to investigate the absolute irreducibility of some special polynomials. Then we use this bound to initiate the induction and derive our main result.

1.3. Main result

For a real $\varepsilon > 0$ we set

$$\eta_1(\varepsilon) = \eta_2(\varepsilon) = \frac{7}{27}\varepsilon, \quad (1.5)$$

and define the sequence $\eta_n(\varepsilon)$, $n = 3, 4, \dots$, recursively as follows

$$\eta_n(\varepsilon) = \frac{7\varepsilon}{18\kappa_n(\varepsilon)}, \quad (1.6)$$

where

$$\kappa_n(\varepsilon) = \left\lceil \frac{n-2-7\varepsilon/3}{2\eta_{n-1}(\varepsilon)} + 3 \right\rceil. \quad (1.7)$$

THEOREM 1.1. *Let $f(X) \in \mathbb{F}_p[X]$ be a polynomial of degree $n \geq 1$, and let $\mathcal{G} \subseteq \mathbb{F}_p^*$ be a subgroup of order $\tau \geq p^{3/7+\varepsilon}$ for some fixed $\varepsilon > 0$. Then*

$$S(\mathcal{G}; f) \ll \tau p^{-\eta_n(\varepsilon)},$$

where $\eta_n(\varepsilon)$ is defined by (1.5) and (1.6).

Remark 1.2. We have included the case of linear polynomials in Theorem 1.1, but as we have mentioned in this case stronger results are available, see, for example, [21].

Remark 1.3. It is obvious from our argument that if some information about the sparsity of the polynomial f in Theorem 1.1 is known, then this can be accommodated in a stronger bound with $\eta_r(\varepsilon)$ instead of $\eta_n(\varepsilon)$ where $r \leq n$ is the number of monomials in f .

Since it may not be easy to understand the behaviour of the sequence $\eta_n(\varepsilon)$ from (1.6) and (1.7), here we give some clarifying examples.

First we compute explicitly,

$$\eta_3(\varepsilon) = \frac{7\varepsilon}{18 \lceil 27\varepsilon^{-1}/14 - 3/2 \rceil}.$$

We also notice that a simple inductive argument shows that for, say, $\varepsilon < 1/2$, for some absolute constant $c > 0$ we have

$$\eta_n(\varepsilon) \geq c \frac{(7\varepsilon/9)^{n-1}}{(n-2)!}.$$

(certainly for $\varepsilon > 1/2$ the Weil bound (1.3) is much stronger).

2. Algebraic geometry background

2.1. Rational points on absolutely irreducible curves

Let q be a prime power. It is well known that by the Weil bound we have

$$\#\{(x, y) \in \mathbb{F}_q^2 : F(x, y) = 0\} = q + O\left(d^2 q^{1/2}\right) \tag{2.1}$$

for any absolutely irreducible polynomial $F(X, Y) \in \mathbb{F}_q[X, Y]$ of degree d (see, for example, [15, section X.5, equation (5.2)]). One can see that (2.1) is a genuine asymptotic formula only for $d = O(q^{1/4})$ and is in fact weaker than the trivial bound

$$\#\{(x, y) \in \mathbb{F}_q^2 : F(x, y) = 0\} = O(dq)$$

for $d \geq q^{1/2}$, which is exactly the range of our interest. To obtain nontrivial bounds for such large values of d we recall the following result, which is a combination of [27, theorem (i)] with the Weil bound (2.1) (and the trivial inequality $p + 2d^2 p^{1/2} \leq 3p$ for $d \leq p^{1/4}$).

LEMMA 2.1. *Let p be prime and let $F(X, Y) \in \mathbb{F}_p[X, Y]$ be an absolutely irreducible polynomial of degree $d < p$. Then*

$$\#\{(x, y) \in \mathbb{F}_p^2 : F(x, y) = 0\} \leq 4d^{4/3} p^{2/3} + 3p.$$

2.2. Absolute irreducibility of some polynomials

To apply the bound of Lemma 2.1 we need to establish absolute irreducibility of polynomials relevant to our applications. We present it in a general form for arbitrary finite fields, as it may be useful for other applications. Below we use a natural mapping of integers into elements of a finite field \mathbb{F}_q of q elements of characteristic p via the reduction modulo p .

LEMMA 2.2. *Given integers $n > m \geq 1$ with $\gcd(m, n) = 1$, there exists a non-zero polynomial $\Delta(U, V) \in \mathbb{Z}[U, V]$, such that for every prime power q and positive integer s with*

$\gcd(s, q) = 1$ and $A, B \in \mathbb{F}_q$ with $\Delta(A, B) \neq 0$, the polynomial

$$F(X, Y) = (X^{sm} + Y^{sm} - A)^n - (X^{sn} + Y^{sn} - B)^m \in \mathbb{F}_q[X, Y]$$

is absolutely irreducible.

Proof. Let us begin by considering the case $s = 1$.

We introduce a new variable Z and note that for $s = 1$ the curve $F = 0$ is isomorphic to

$$\begin{cases} X^m + Y^m - A = Z^m \\ X^n + Y^n - B = Z^n. \end{cases} \quad (2.2)$$

The isomorphism is the projection to the X, Y -plane, with inverse given by

$$(X, Y) \mapsto (X, Y, (X^m + Y^m - A)^u (X^n + Y^n - B)^v)$$

with some fixed integers u and v satisfying

$$mu + nv = 1.$$

The two equations in (2.2) have gradients

$$m(X^{m-1}, Y^{m-1}, -Z^{m-1}) \quad \text{and} \quad n(X^{n-1}, Y^{n-1}, -Z^{n-1}),$$

respectively. For the curve defined by the system (2.2) to be singular at a point (x, y, z) the corresponding gradients have to be linearly dependent. This condition, when fed back into (2.2) gives a relation between A, B .

More explicitly we proceed as follows.

First, we seek the polynomial Δ in the form

$$\Delta = mn\Delta_0 \quad (2.3)$$

with some $\Delta_0(U, V) \in \mathbb{Z}[U, V]$, thus $\Delta(A, B) \neq 0$, guarantees that $mn \neq 0$ in \mathbb{F}_q .

If (x, y, z) is a solution to (2.2) with $x = y = 0$, then $z^m = -A$, $z^n = -B$, and so $(-A)^n = (-B)^m$. Thus we also request

$$((-A)^n - (-B)^m) \mid \Delta(A, B). \quad (2.4)$$

The possibilities $x = z = 0$ and $y = z = 0$ can be similarly treated and lead to the requirement

$$(A^n - B^m) \mid \Delta(A, B). \quad (2.5)$$

If $x = 0$ and $yz \neq 0$, then the gradient condition gives $y = \zeta z$ with some $\zeta^{n-m} = 1$ and (2.2) gives

$$(\zeta^m - 1)z^m = A \quad \text{and} \quad (\zeta^n - 1)z^n = B.$$

Therefore, $(\zeta^n - 1)^m A^n = (\zeta^m - 1)^n B^m$, and we also request

$$\prod_{\zeta^{n-m}=1} ((\zeta^n - 1)^m A^n - (\zeta^m - 1)^n B^m) \mid \Delta(A, B), \quad (2.6)$$

where the product is taken over all roots of unity ζ with $\zeta^{n-m} = 1$.

If $xyz \neq 0$, then we see from (2.7) that there has to be a constant λ with

$$x^{n-m} = y^{n-m} = z^{n-m} = \lambda, \tag{2.7}$$

so we can write $y = \zeta_1 x$, and $z = \zeta_2 x$ with $\zeta_1^{n-m} = \zeta_2^{n-m} = 1$, leading to

$$\prod_{\substack{\zeta_1^{n-m}=1 \\ \zeta_2^{n-m}=1}} ((1 + \zeta_1^n - \zeta_2^n)^m A^n - (1 + \zeta_1^m - \zeta_2^m)^n B^m) \mid \Delta(A, B), \tag{2.8}$$

where the product is taken over all pairs of roots of unity (ζ_1, ζ_2) with $\zeta_1^{n-m} = \zeta_2^{n-m} = 1$.

A similar argument works at infinity and shows that, for generic A and B , the curve is smooth.

Given X, Y , there is a unique choice of Z satisfying (2.2), so the projection to the X, Y does not acquire singularities from distinct points in three-space. The only singularities are cusps coming from a vertical tangent line which are unbranched (since the curve in three-space is smooth). However, a reducible plane curve has singular points with more than one branch wherever two components meet. Hence (2.2) is an irreducible curve. (See [12, chapter 16] for a detailed exposition of branches of curve singularities.)

We have shown that, for $s = 1$, the polynomial F is absolutely irreducible. We consider the algebraic curve \mathcal{C} which is a non-singular projective model of $F = 0$ (still with $s = 1$).

Suppose $n > m > 1$. We now use an argument similar to [19, lemma 4.3].

For a point $P = (0, y_0)$ on the curve $F = 0$ we have

$$\frac{\partial F}{\partial X}(0, y_0) = 0.$$

Next we show that the point $P = (0, y_0)$ is a simple point on the curve $F = 0$ with

$$\frac{\partial F}{\partial Y}(0, y_0) \neq 0 \tag{2.9}$$

(for generic A and B). It now suffices to show that the discriminant $D(A, B)$ of $F(0, Y) \in \mathbb{F}_q[Y]$ (as a polynomial in A, B) is not identically zero, and can be chosen with integer coefficients which depend only on m and n .

Taking $A = 1$ and $B = 0$, the polynomial F specialises to

$$(Y^m - 1)^n - Y^{mn} = \prod_{\xi^n=1} (Y^m(1 - \xi) - 1),$$

which has a non-zero discriminant as each factor $(Y^m(1 - \xi) - 1)$ is square-free and these factors are relatively prime. Thus we impose the condition

$$D(A, B) \mid \Delta(A, B). \tag{2.10}$$

It remains to choose $\Delta(A, B) \in \mathbb{Z}[A, B]$ as an arbitrary fixed polynomial which depends only on m and n and satisfies the divisibility conditions (2.3), (2.4), (2.5), (2.6), (2.8) and (2.10).

We now consider the case of arbitrary $s \geq 1$ (and $n > m > 1$).

So P corresponds to a place of \mathcal{C} . We consider the functions x, y on \mathcal{C} that satisfy the equation $F(x, y) = 0$. The function x has a simple zero at P , hence is not a power of another function on \mathcal{C} . It follows from [26, proposition 3.7.3], that the equation $U^s = x$ is irreducible

over the function field of \mathcal{C} and defines a cover \mathcal{D} of \mathcal{C} . Now, consider any point Q on \mathcal{D} above a point $(x_0, 0)$ on \mathcal{C} . Since x is not zero at $(x_0, 0)$ (for generic A, B), the curve \mathcal{D} is locally isomorphic to \mathcal{C} near Q and we conclude, as above, that the function y on \mathcal{D} has a simple zero at Q and, in particular, is not a power of another function on \mathcal{D} . Again, we conclude that the equation $W^s = y$ is irreducible over the function field of \mathcal{D} and defines a cover \mathcal{E} of \mathcal{D} . In other words, $F(U^s, W^s) = 0$ is an absolutely irreducible equation defining the curve \mathcal{E} , which concludes the case $n > m > 1$.

We are now left with $n > m = 1$. It is still true that a point $P = (0, y_0)$ is a simple point on the curve $F = 0$ with (2.9) (for generic A, B). Indeed, if

$$0 = \frac{\partial F}{\partial Y}(0, y_0) = n(y_0^{n-1} - (y_0 - B)^{n-1}),$$

then combining this with

$$0 = F(0, y_0) = -A + y_0^n - (y_0 - B)^n,$$

we derive

$$0 = -A + y_0^n - (y_0 - B)^n = -A + y_0^n - (y_0 - B)y_0^{n-1} = -A + By_0^{n-1}.$$

Hence

$$0 = By_0^{n-1} - B(y_0 - B)^{n-1} = A - B(y_0 - B)^{n-1}.$$

Considering the resultant of $B(Y - B)^{n-1} - A$ and $BY^{n-1} - A$ (which clear does not vanish for $A = 1$ and $B = 0$ and thus is a nontrivial polynomial) gives a contradiction for generic A, B . The proof then continues as before in the case $n > m > 1$.

3. Exponential sums and systems of diagonal equations

3.1. Exponential sums and the number of solutions to some systems of equations

Here we collect some previous results on exponential sums and also about links between these bounds and the number of solutions to some congruences.

Given an integer vector $\mathbf{n} = (n_1, \dots, n_r) \in \mathbb{Z}^r$ with $n_r > \dots > n_1 \geq 1$, and a subgroup $\mathcal{G} \subseteq \mathbb{F}_p^*$, we denote by $Q_k(\mathbf{n}; \mathcal{G})$ the number of solutions to the following system of r equations

$$g_1^{n_i} + \dots + g_k^{n_i} = g_{k+1}^{n_i} + \dots + g_{2k}^{n_i}, \quad i = 1, \dots, r, \tag{3.1}$$

$$g_1, \dots, g_{2k} \in \mathcal{G}.$$

The following link between $S(\mathcal{G}; f)$ and $Q_k(\mathbf{n}; \mathcal{G})$ is a slight variation of several previous results of a similar spirit.

LEMMA 3.1. *Let*

$$f(X) = a_r X^{n_r} + \dots + a_1 X^{n_1} \in \mathbb{F}_p[X]$$

with nonzero coefficients $a_1, \dots, a_r \in \mathbb{F}_p^$ and integer exponents $n_r > \dots > n_1 \geq 1$. Then, for a subgroup $\mathcal{G} \subseteq \mathbb{F}_p^*$ and for any positive integers k and ℓ , we have*

$$|S(\mathcal{G}; f)|^{2k\ell} \leq p^r \tau^{2k\ell - 2k - 2\ell} Q_k(\mathbf{n}; \mathcal{G}) Q_\ell(\mathbf{n}; \mathcal{G}).$$

Proof. We start by noticing that, for any $h \in \mathcal{G}$, we have

$$S(\mathcal{G}; f) = \sum_{g \in \mathcal{G}} \mathbf{e}_p (a_1(hg)^{n_1} + \dots + a_r(hg)^{n_r}).$$

Hence, for any integer $k \geq 1$ we have

$$\begin{aligned} \tau (S(\mathcal{G}; f))^k &= \sum_{h \in \mathcal{G}} \left(\sum_{g \in \mathcal{G}} \mathbf{e}_p (a_1(hg)^{n_1} + \dots + a_r(hg)^{n_r}) \right)^k \\ &= \sum_{h \in \mathcal{G}} \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r) \mathbf{e}_p (\lambda_1 h^{n_1} + \dots + \lambda_r h^{n_r}), \end{aligned}$$

where $J_k(\lambda_1, \dots, \lambda_r)$ is the number of solutions to the following system of equations:

$$\begin{aligned} a_i (g_1^{n_i} + \dots + g_k^{n_i}) &= \lambda_i, \quad i = 1, \dots, r, \\ g_1, \dots, g_k &\in \mathcal{G}. \end{aligned}$$

Hence, changing the order of summations, we obtain

$$\tau |S(\mathcal{G}; f)|^k \leq \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r) \left| \sum_{h \in \mathcal{G}} \mathbf{e}_p (\lambda_1 h^{n_1} + \dots + \lambda_r h^{n_r}) \right|.$$

Observe that since $a_1, \dots, a_r \neq 0$, we have

$$\begin{aligned} \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r) &= \tau^k, \\ \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r)^2 &= Q_k(\mathbf{n}; G), \end{aligned}$$

where $\mathbf{n} = (n_1, \dots, n_r)$.

Writing

$$J_k(\lambda_1, \dots, \lambda_r) = J_k(\lambda_1, \dots, \lambda_r)^{1-1/\ell} \left(J_k(\lambda_1, \dots, \lambda_r)^2 \right)^{1/2\ell}$$

and applying the Hölder inequality, we derive

$$\begin{aligned} \tau^{2\ell} (S(\mathcal{G}; f))^{2k\ell} &\leq \left(\sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r) \right)^{2\ell-2} \\ &\quad \times \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} J_k(\lambda_1, \dots, \lambda_r)^2 \\ &\quad \times \sum_{\lambda_1, \dots, \lambda_r \in \mathbb{F}_p} \left| \sum_{h \in \mathcal{G}} \mathbf{e}_p (\lambda_1 h^{n_1} + \dots + \lambda_r h^{n_r}) \right|^{2\ell} \\ &= p^r \tau^{k(2\ell-2)} Q_k(\mathbf{n}; G) Q_\ell(\mathbf{n}; G) \end{aligned}$$

which concludes the proof.

We now establish a link in the opposite direction, that is, from bounds on exponential sums to bounds on $Q_k(\mathbf{n}; \mathcal{G})$.

LEMMA 3.2. *Let $r \geq 2$ and let $\varepsilon \geq 0$ be fixed. Assume that there is some fixed $\eta > 0$ (depending only on r and ε) such that for all nonzero vectors $(a_1, \dots, a_r) \in \mathbb{F}_p^r$ and for a vector $\mathbf{n} = (n_1, \dots, n_r) \in \mathbb{Z}^r$ with $n_r > \dots > n_1 \geq 1$, for a subgroup $\mathcal{G} \subseteq \mathbb{F}_p^*$ of order τ with*

$$p^{3/7+\varepsilon} \leq \tau \leq p^{3/4} \tag{3.2}$$

we have

$$\sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \ll \tau p^{-\eta}. \tag{3.3}$$

Then for any integer $k \geq 3$ we have

$$Q_k(\mathbf{n}; \mathcal{G}) \ll \tau^{2k} p^{-\xi},$$

where $\xi = \min\{r, \eta(2k - 6) + 1 + 7\varepsilon/3\}$.

Proof. Using the orthogonality of characters, we write

$$\begin{aligned} Q_k(\mathbf{n}, \mathcal{G}) &= \frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k} \\ &= \frac{\tau^{2k}}{p^r} + \frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k}. \end{aligned} \tag{3.4}$$

Now, using our assumption (3.3) we obtain

$$\begin{aligned} \frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k} \\ \ll \frac{(\tau p^{-\eta})^{2k-6}}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^6. \end{aligned}$$

Dropping the restriction $(a_1, \dots, a_r) \neq \mathbf{0}$ from the summation, we now obtain

$$\begin{aligned} \frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k} \\ \ll \frac{(\tau p^{-\eta})^{2k-6}}{p^r} \sum_{a_1, \dots, a_r \in \mathbb{F}_p} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^6 \end{aligned}$$

$$= (\tau p^{-\eta})^{2k-6} Q_3(\mathbf{n}; \mathcal{G}).$$

Since $r \geq 2$, we obviously have

$$Q_3(\mathbf{n}; \mathcal{G}) \leq Q_3(n_1, n_2; \mathcal{G}).$$

Thus applying Corollary 3.4 in Section 3.2 below and using that under our assumption (3.2) we have $\tau^{11/3} \geq \tau^5/p$, we obtain

$$\begin{aligned} \frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k} &\ll (\tau p^{-\eta})^{2k-6} \tau^{11/3} \\ &= \tau^{2k-7/3} p^{-\eta(2k-6)}. \end{aligned}$$

Recalling (3.2) again we see that

$$\frac{1}{p^r} \sum_{\substack{a_1, \dots, a_r \in \mathbb{F}_p \\ (a_1, \dots, a_r) \neq \mathbf{0}}} \left| \sum_{g \in \mathcal{G}} \mathbf{e}_p(a_1 g^{n_1} + \dots + a_r g^{n_r}) \right|^{2k} \ll \tau^{2k} p^{-\eta(2k-6)-1-7\epsilon/3},$$

which together with (3.4) concludes the proof.

3.2. *Bounds on the number of solutions to some systems of equations in six variables*

We start with an observation that the results of this section are independent of those in Section 3.1 and hence there is no logical problem in our use of them in the proof of Lemma 3.2.

For $r = 2$ and $\mathbf{n} = (m, n)$ we write $Q_k(m, n; \mathcal{G})$ for $Q_k(\mathbf{n}; \mathcal{G})$.

Here we obtain some bounds on $Q_3(m, n; \mathcal{G})$. In fact, it is easier to work with the following system of equations:

$$\begin{cases} x_1^{sm} + x_2^{sm} + x_3^{sm} = x_4^{sm} + x_5^{sm} + x_6^{sm} \\ x_1^{sn} + x_2^{sn} + x_3^{sn} = x_4^{sn} + x_5^{sn} + x_6^{sn} \end{cases}, \quad x_1, \dots, x_6 \in \mathbb{F}_p^*, \tag{3.5}$$

instead of the system of the type (3.1) with group elements.

Denoting by $T_3(m, n; s)$ the number of solutions to (3.5) we see that

$$Q_3(m, n; \mathcal{G}) = s^{-6} T_3(m, n; s), \tag{3.6}$$

where

$$s = \frac{p-1}{\tau}$$

and, as before, $\tau = \#\mathcal{G}$.

LEMMA 3.3. *For integers $n > m > 0$, we have*

$$T_3(m, n; s) \ll s^{7/3} p^{11/3} + sp^4.$$

Proof. First we note that if $\gcd(m, n) = d$ then

$$T_3(m, n; s) \leq e^6 T_3(m/d, n/d; s),$$

where $e = \gcd(d, p - 1)$.

Hence we can assume that

$$\gcd(m, n) = 1, \tag{3.7}$$

which enables us to apply Lemma 2.2.

We now fix x_4, x_5 and x_6 and thus we obtain $(p - 1)^3$ systems of equations of the form

$$\begin{cases} x_1^{sm} + x_2^{sm} + x_3^{sm} = A \\ x_1^{sn} + x_2^{sn} + x_3^{sn} = B \end{cases}, \quad x_1, x_2, x_3 \in \mathbb{F}_p^*,$$

where

$$A = x_4^{sm} + x_5^{sm} + x_6^{sm} \quad \text{and} \quad B = x_4^{sn} + x_5^{sn} + x_6^{sn},$$

from which we derive

$$(x_1^{sm} + x_2^{sm} - A)^n = (x_1^{sn} + x_2^{sn} - B)^m. \tag{3.8}$$

Under the assumption (3.7), let the polynomials $\Delta(U, V) \in \mathbb{Z}[U, V]$ be as in Lemma 2.2.

Since Δ depends only on m and n , we see that if p is large enough, Δ is a non-zero polynomial modulo p .

We assume first that $\Delta(A, B) = 0$ for a pair $(A, B) \in \mathbb{F}_p^2$ as above. Thus

$$\Delta(x_4^{sm} + x_5^{sm} + x_6^{sm}, x_4^{sn} + x_5^{sn} + x_6^{sn}) = 0. \tag{3.9}$$

If $\Delta(X^{sm} + x_5^{sm} + x_6^{sm}, X^{sn} + x_5^{sn} + x_6^{sn})$, as a polynomial in X , is not identically zero for some $(x_5, x_6) \in \mathbb{F}_p^2$, then obviously it has $O(s)$ zeros. Thus, in this case, the equation (3.9) has $O(sp^2)$ solutions $(x_4, x_5, x_6) \in \mathbb{F}_p^3$.

On the other hand, if $\Delta(X^{sm} + x_5^{sm} + x_6^{sm}, X^{sn} + x_5^{sn} + x_6^{sn})$, as a polynomial in X , is identically zero, then it also holds for $X = 0$, thus

$$\Delta(x_5^{sm} + x_6^{sm}, x_5^{sn} + x_6^{sn}) = 0. \tag{3.10}$$

Now, a similar argument shows that (3.10) holds for $O(sp)$ pairs (x_5, x_6) for which there are at most p values of x_4 .

Therefore, the equation (3.9) has $O(sp^2)$ solutions $(x_4, x_5, x_6) \in \mathbb{F}_p^3$ in total.

For each of such $O(sp^2)$ values of (x_4, x_5, x_6) the corresponding equation (3.8) is nontrivial since it contains a unique term $nx_1^{sm(n-1)}x_2^{sm}$ and hence has $O(sp)$ solutions (x_1, x_2) after which there are $O(s)$ possible values for x_3 (we recall that the implied constants may depend on n). Hence, the total contribution from the case $\Delta(A, B) = 0$ is $O(s^3p^3)$.

If $\Delta(A, B) \neq 0$, then for the corresponding $O(p^3)$ possibilities for $(x_4, x_5, x_6) \in \mathbb{F}_p^3$, by Lemma 2.2, we can apply Lemma 2.1 to bound the number of solutions to (3.8) (after which we have $O(s)$ possibilities for x_3). Hence, the total contribution from the case $\Delta(A, B) \neq 0$ is $O(s(s^{4/3}p^{2/3} + p)p^3)$.

Therefore $T_3(m, n; s) \leq s^3 p^3 + s^{7/3} p^{11/3} + s p^4$. Since $s^3 p^3 \leq s^{7/3} p^{11/3}$ for $s \leq p$, the result follows.

Recalling (3.6), we see that Lemma 3.3 implies the following.

COROLLARY 3.4. *For integers $n > m > 0$, we have*

$$Q_3(m, n; \mathcal{G}) \ll \tau^{11/3} + \tau^5/p.$$

Remark 3.5. We recall that the method of Kurlberg and Rudnick [13, lemma 5], immediately implies that $Q_2(m, n; \mathcal{G}) \ll \tau^2$. However this bound is not sufficient for our purpose.

3.3. *Bounds on monomial and binomial sums*

First we recall the following result of Shkredov [21, theorem 1] (with a slight generalisation and also combined with a direct implication of (1.1)).

LEMMA 3.6. *Let $f(X) = aX^n \in \mathbb{F}_p[X]$ of degree $n \geq 1$ and with $a \neq 0$, and let $\mathcal{G} \subseteq \mathbb{F}_p^*$ be a subgroup \mathcal{G} of order τ . Then*

$$S(\mathcal{G}; f) \ll \min\{p^{1/2}, \tau^{1/2} p^{1/6} (\log p)^{1/6}\}.$$

Proof. We remark that the result of Shkredov [21, theorem 1] corresponds to $n = 1$. Otherwise we note that

$$S(\mathcal{G}; f) = d \sum_{x \in \mathcal{G}^d} \mathbf{e}_p(ax),$$

where $d = \gcd(\tau, n)$ and $\mathcal{G}^d = \{g^d : g \in \mathcal{G}\}$.

We now derive the following estimate, which improves (1.3) for $\tau \leq p^{21/40}$, remains non-trivial for $\tau \geq p^{3/7+\varepsilon}$ for any fixed $\varepsilon > 0$ and which we believe is of independent interest. For this, we apply Lemma 3.1 with $k = \ell = 3$ and we use Corollary 3.4.

LEMMA 3.7. *Let $f(X) = aX^m + bX^n \in \mathbb{F}_p[X]$ with integers $n > m \geq 1$ and $(a, b) \neq (0, 0)$, and let $\mathcal{G} \subseteq \mathbb{F}_p^*$ be a subgroup of order τ . Then*

$$S(\mathcal{G}; f) \ll \tau^{20/27} p^{1/9}.$$

Proof. If $ab = 0$ then the result is instant from Lemma 3.6.

Hence we now assume $a, b \in \mathbb{F}_p^*$. We apply Lemma 3.1 with

$$(k, \ell) = (3, 3)$$

and the bound of Corollary 3.4. We also note that for $\tau > p^{21/40}$ we have

$$p^{1/2} \leq \tau^{20/27} p^{1/9}.$$

Hence we only need to apply Corollary 3.4 for $\tau \leq p^{21/40}$ in which case $\tau^{11/3} \geq \tau^5/p$, and thus in this case we simply have $Q_3(m, n; \mathcal{G}) \ll \tau^{11/3}$ and the result follows.

4. Proof of Theorem 1.1

4.1. Preliminaries and the basis of induction

We prove the result by induction on the number of terms r in the polynomial

$$f(X) = a_r X^{n_r} + \dots + a_1 X^{n_1} \in \mathbb{F}_p[X]$$

with nonzero coefficients $a_1, \dots, a_r \in \mathbb{F}_p^*$ and integer exponents $n = n_r > \dots > n_1 \geq 1$.

We see from Lemma 3.7 that for $r = 1, 2$ the condition (3.3) of Lemma 3.2 is satisfied with

$$\eta = \eta_1(\varepsilon) \quad \text{and} \quad \eta = \eta_2(\varepsilon),$$

respectively, where $\eta_1(\varepsilon)$ and $\eta_2(\varepsilon)$ are given by (1.5), which form the basis of induction.

4.2. Inductive step

Assume that the result holds for all nontrivial polynomials of degree at most n with at most $r - 1$ monomials and we prove it for polynomials with $r \geq 3$ monomials. First we note that we can assume that $\tau \leq p^{3/4}$ since otherwise the bound (1.3) is stronger than that of Theorem 1.1.

In particular, the condition (3.2) of Lemma 3.2 is satisfied. We fix some arbitrary positive integers k, ℓ, u and v with $u, v \leq r$. Let

$$\mathbf{n}_u = (n_1, \dots, n_u) \quad \text{and} \quad \mathbf{n}_v = (n_1, \dots, n_v).$$

We now use the trivial bounds

$$Q_k(\mathbf{n}; \mathcal{G}) \leq Q_k(\mathbf{n}_u; \mathcal{G}) \quad \text{and} \quad Q_\ell(\mathbf{n}; \mathcal{G}) \leq Q_\ell(\mathbf{n}_v; \mathcal{G}). \tag{4.1}$$

In fact, we choose $u = r - 1$ and $v = 2$. Furthermore, we recall the definition (1.7) and set

$$k = \kappa_r(\varepsilon) \quad \text{and} \quad \ell = 3, \tag{4.2}$$

in which case, using the induction assumption, by Lemma 3.2, used with $u = r - 1$ instead of r , we have

$$Q_k(\mathbf{n}_{r-1}; \mathcal{G}) \ll \tau^{2k} / p^{r-1}, \tag{4.3}$$

while by Corollary 3.4, using that $\tau \leq p^{3/4}$, we obtain

$$Q_3(\mathbf{n}_2; \mathcal{G}) \ll \tau^{11/3} + \tau^5 / p \ll \tau^{11/3}. \tag{4.4}$$

Indeed, (4.3) follows from the definition of $\kappa_r(\varepsilon)$ in (1.7), which ensures that $\xi = r - 1$ in Lemma 3.2.

Next, substituting the bounds (4.1), (4.3) and (4.4) in the estimate of Lemma 3.1, we obtain

$$S(\mathcal{G}; f)^{6k} \ll p^r \tau^{4k-6} \frac{\tau^{2k}}{p^{r-1}} \tau^{11/3} = \tau^{6k-7/3} p \leq \tau^{6k-7\varepsilon/3}.$$

Hence

$$S(\mathcal{G}; f) \ll \tau^{1-7\varepsilon/18k}.$$

Recalling the definition (1.6) and the choice of k in (4.2), we conclude the proof.

5. Comments

If ϑ is a generator of \mathcal{G} then the sum $S(\mathcal{G}; f)$ can be written as

$$S(\mathcal{G}; f) = \sum_{x=1}^{\tau} \mathbf{e}_p(f(\vartheta^x)).$$

This reformulation also allows us to generalise these sums to twisted sums,

$$S_b(\mathcal{G}; f) = \sum_{x=1}^{\tau} \mathbf{e}_p(f(\vartheta^x)) \exp(2\pi ibx/\tau),$$

to which all our results apply without any changes (with just minor typographic adjustments). In turn, together with the well-known completing technique (see, for example, [8, section 12.2]) bounds on the sums $S_b(\mathcal{G}; f)$ lead to bounds on incomplete sums

$$\sum_{x=1}^N \mathbf{e}_p(f(\vartheta^x)), \quad 1 \leq N \leq \tau.$$

We note that in [18] sequences of the form $(f(\vartheta^x))$ have been studied as sources of pseudorandom numbers, but with nontrivial results only in the case of periods $\tau > p^{1/2+\varepsilon}$, while our results allows us to extend this range to $\tau > p^{3/7+\varepsilon}$.

We also note that in [16, 29] the sequence

$$(a\vartheta^x + b)^{-1}, \quad n = 1, 2, \dots,$$

has been suggested as a source of pseudorandom numbers. Unfortunately neither the method of Bourgain [1] nor of this work applies to the corresponding exponential sums

$$\sum_{x=1}^N \mathbf{e}_p\left((a\vartheta^x + b)^{-1}\right), \quad 1 \leq N \leq \tau,$$

(with a natural convention that the values with $a\vartheta^x = -b$ are excluded), which are necessary for investigating this sequence. So we leave a question of obtaining such nontrivial bounds for $\tau < p^{1/2}$ as an open problem. Even the case of complete sums

$$\sum_{g \in \mathcal{G}} \mathbf{e}_p\left((ag + b)^{-1}\right)$$

is of interest.

Acknowledgements. During the preparation of this work, the first two authors (A.O. and I.E.S.) were partially supported by the Australian Research Council Grant DP200100355. The third author (J.F.V.) was partially supported by a grant from the Ministry of Business, Innovation and Employment of New Zealand.

REFERENCES

- [1] J. BOURGAIN. Mordell's exponential sum estimate revisited. *J. Amer. Math. Soc.* **18** (2005), 477–499.
- [2] J. BOURGAIN. Multilinear exponential sums in prime fields under optimal entropy condition on the sources. *Geom. and Funct. Anal.* **18** (2009), 1477–1502.
- [3] J. BOURGAIN. On the Vinogradov mean value. *Proc. Steklov Math. Inst.* **296** (2017), 30–40.

- [4] J. BOURGAIN, A. A. GLIBICHUK and S. V. KONYAGIN. Estimates for the number of sums and products and for exponential sums in fields of prime order. *J. London Math. Soc.* **73** (2006), 380–398.
- [5] D. DI BENEDETTO, M. Z. GARAEV, V. C. GARCIA, D. GONZALEZ-SANCHEZ, I. E. SHPARLINSKI and C. A. TRUJILLO. New estimates for exponential sums over multiplicative subgroups and intervals in prime fields. *J. Number Theory*, **215** (2020), 261–274.
- [6] M. Z. GARAEV. Sums and products of sets and estimates of rational trigonometric sums in fields of prime order. *Russian Math. Surveys* **65** (2010), 599–658 (transl. from *Uspekhi Mat. Nauk*).
- [7] D. R. HEATH-BROWN and S. V. KONYAGIN. New bounds for Gauss sums derived from k th powers, and for Heilbronn’s exponential sum. *Quart. J. Math.* **51** (2000), 221–235.
- [8] H. IWANIEC and E. KOWALSKI. *Analytic Number Theory* (Amer. Math. Soc., Providence, RI, 2004).
- [9] B. KERR. Incomplete Gauss sums modulo primes. *Quart. J. Math.* **69** (2018), 729–745.
- [10] B. KERR and S. MACOURT. Multilinear exponential sums with a general class of weights. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **22** (2021), 1105–1130.
- [11] S. V. KONYAGIN. *Bounds of exponential sums over subgroups and Gauss sums*. Proc 4th Intern. Conf. Modern Problems of Number Theory and Appl. (Moscow Lomonosov State Univ., Moscow, 2002), 86–114 (in Russian).
- [12] E. KUNZ. *Introduction to Plane Algebraic Curves*. (Birkhauser, 2005).
- [13] P. KURLBERG and Z. RUDNICK. On quantum ergodicity for linear maps of the torus. *Comm. Math. Phys.* **222** (2001), 201–227.
- [14] W.-C. W. LI. *Number Theory with Applications* (World Scientific, Singapore, 1996).
- [15] D. LORENZINI. *An Invitation to Arithmetic Geometry*, (Amer. Math. Soc., Providence, RI, 1996).
- [16] W. MEIDL and A. WINTERHOF. On the linear complexity profile of some new explicit inverse pseudorandom numbers. *J. Complexity*. **20** (2004) 350–355.
- [17] C. J. MORENO and O. MORENO. Exponential sums and Goppa codes. I. *Proc. Amer. Math. Soc.* **111** (1991), 523–531.
- [18] H. NIEDERREITER and A. WINTERHOFF. On the distribution of some new explicit nonlinear congruential pseudorandom numbers. *Proc. SETA 2004*. Lecture Notes in Comput. Sci. vol. 3486 (Springer, 2005), pp. 266–274.
- [19] A. OSTAFE, I. E. SHPARLINSKI and J. F. VOLOCH. Equations and character sums with matrix powers, Kloosterman sums over small subgroups and quantum ergodicity. *Internat. Math. Res. Notices* (to appear).
- [20] S. N. POPOVA. On sums of products in $\mathbb{F}_p \times \mathbb{F}_p$. *Mat. Zametki* **106** (2019), 262–279 (in Russian).
- [21] I. D. SHKREDOV. On exponential sums over multiplicative subgroups of medium size. *Finite Fields and Appl.* **30** (2014), 72–87.
- [22] I. D. SHKREDOV. Some remarks on the asymmetric sum-product phenomenon. *Moscow J. Combin. and Number Theory* **8** (2019), 15–41.
- [23] I. E. SHPARLINSKI. Estimates for Gauss sums. *Math. Notes*. **50** (1991), 740–746 (transl. from *Mat. Zametki*).
- [24] I. E. SHPARLINSKI and J. F. VOLOCH. Binomial exponential sums. *Ann. Scuola Norm. Sup. Pisa Cl. Sci.* **XXI** (2020), 931–941.
- [25] I. E. SHPARLINSKI and Q. WANG. Exponential sums with sparse polynomials over finite fields. *SIAM J. Discrete Math.* **35** (2021), 976–987.
- [26] H. STICHTENOTH. *Algebraic Function Fields and Codes* (Springer-Verlag, Berlin, 2009).
- [27] J. F. VOLOCH. On the number of values taken by a polynomial over a finite field. *Acta Arith.*, **52** (1989), 197–201.
- [28] A. WEIL. On some exponential sums. *Proc. Nat. Sci. Acad. Sci U.S.A.* **34** (1948), 204–207.
- [29] A. WINTERHOF. On the distribution of some new explicit inverse pseudorandom numbers and vectors. *Monte Carlo and Quasi-Monte Carlo Methods 2004*. (Springer, 2006), pp. 487–499.