EXPONENTIAL SUMS OVER SMALL SUBGROUPS, REVISITED

EMMANUEL KOWALSKI

Abstract. This is an expository account of the proof of the theorem of Bourgain, Glibichuk and Konyagin which provides non-trivial bounds for exponential sums over very small multiplicative subgroups of prime finite fields.

... this peaking of the whale's flukes is perhaps the grandest sight to be seen in all animated nature,
H. Melville, *Moby-Dick*, Ch. lxxxvi.

1. Introduction

In the theory of exponential sums in number theory, the study of "short" sums remains one of the most mysterious. Truly robust methods, suitable for the variety of sums that appear in applications, are lacking in many cases.

This note is an exposition of the proof by Bourgain, Glibichuk and Konyagin of a remarkable estimate of this kind. It concerns exponential sums over "small" subgroups of $\mathbf{F}_p^\times$, and is especially noteworthy for the techniques, based on additive combinatorics, which enter into the proof.

The precise result is the following:

Theorem 1.1 (Bourgain, Glibichuk and Konyagin)**.** *Let $\gamma > 0$ be a real number. There exists a real number $\nu > 0$, depending only on γ, such that for any prime number p and any subgroup $\mathrm{H} \subset \mathbf{F}_p^\times$ with $|\mathrm{H}| \geqslant p^\gamma$, we have*

$$\sum_{x\in \mathrm{H}} e\Big(\frac{ax}{p}\Big) \ll |\mathrm{H}|p^{-\nu}$$

for any $a \in \mathbf{F}_p^\times$, where the implied constant depends only on γ.

Theorem 1.1 has an equivalent formulation in terms of Gauss sums

$$\mathrm{G}_d(a;p) = \sum_{x\in \mathbf{F}_p} e\Big(\frac{ax^d}{p}\Big)$$

with exponent $d \mid p-1$. Indeed, considering the subgroup

$$\mathrm{H}_d = \{x^d \mid x \in \mathbf{F}_p^\times\}$$

2010 *Mathematics Subject Classification.* 11L07, 11T23.

Key words and phrases. Exponential sums, additive combinatorics, sum-product phenomenon, Balog–Szemerédi–Gowers Theorem, multiplicative energy, random walks on finite abelian groups.

of order $(p-1)/d$, we have

$$\mathrm{G}_d(a;p) = 1 + \sum_{x\in\mathbf{F}_p^\times} e\Big(\frac{ax^d}{p}\Big) = 1 + d\sum_{y\in \mathrm{H}_d} e\Big(\frac{ay}{p}\Big) = 1 + \frac{p-1}{|\mathrm{H}_d|}\sum_{y\in\mathrm{H}} e\Big(\frac{ay}{p}\Big)$$

since each $y \in \mathrm{H}_d$ is of the form $y = x^d$ for d different values of $x \in \mathbf{F}_p^\times$. Hence we see that the estimate of the theorem is equivalent to the bound $\mathrm{G}_d(a;p) \ll p^{1-\nu}$, valid provided $d \leqslant (p-1)p^{-\gamma}$ for some $\gamma > 0$.

Similarly, let H be a subgroup of $\mathbf{F}_p^\times$. We can write

$$\sum_{y\in\mathrm{H}} e\Big(\frac{ay}{p}\Big) = \frac{|\mathrm{H}|}{p-1}\sum_{\mathrm{H}\subset\ker(\chi)}\sum_{y\in\mathbf{F}_p}\chi(y)e\Big(\frac{ay}{p}\Big),$$

where χ runs over the subgroup of characters trivial on H (which has order $(p-1)/|\mathrm{H}|$); using the fact that Gauss sums for non-trivial characters have modulus $\sqrt{p}$, we see that the sums in Theorem 1.1 have modulus at most $\sqrt{p}$. This is non-trivial for $|\mathrm{H}|$ a bit larger than $\sqrt{p}$. (See Remark 5.2, (3) for a different proof of this which does not use Gauss sums.)

Remark 1.2. (1) Using similar methods in combination with significant other ingredients, a number of generalizations of this bound have been obtained, among which we single out the result of Bourgain [3] where non-trivial estimates are obtained for the sums

$$\sum_{x\in\mathbf{F}_p^\times} e\Big(\frac{f(x)}{p}\Big)$$

for $f \in \mathbf{Z}[\mathrm{X}]$ of possibly very large degree, provided the degrees of the non-zero monomials appearing in f satisfy suitable conditions relative to p.

We focus on Theorem 1.1 for definiteness and clarity.

(2) One can wonder about even smaller subgroups, but some restriction is certainly needed since H could be of bounded order. For instance, if p is odd, there is always a subgroup of order 2, namely $\{-1, 1\}$, for which the behavior of the sums is quite clearly rather different.

It would be interesting to see if one could say something interesting for subgroups H of size $\asymp (\log p)^{\mathrm{C}}$ for some constant $\mathrm{C} > 0$.

(3) The dependency of the exponent ν on γ can be made explicit in Theorem 1.1; currently the sharpest result (whose proof involves new ideas) is due to Shkredov [11, Cor. 16].

Remark 1.3. Some of the motivation, generalizations and applications of Theorem 1.1 are discussed in a talk at IAS by Bourgain in December 2008, which is available online [2].

P. Kurlberg [9] has already written a detailed account of the proof of Theorem 1.1, from which we benefited a lot. The first version of the present text was written as part of lecture notes for an introductory course on additive combinatorics taught in the Fall Semester 2023 at ETH Zürich (see [8] for the current draft), but the current presentation is also quite different from that.

Some of the changes we make in comparison with the original paper of Bourgain, Glibichuk and Konyagin (and with Kurlberg's account) are the following:

– The argument, which was originally phrased in terms of probability measures on $\mathbf{F}_p$ is presented in probabilistic language. At least for some readers (starting from the author), this focus brings some additional insights and intuition.

– In addition, we order and phrase the main steps of the proof rather differently (compare Proposition 5.1 with [9, Prop. 3.1], for instance; these are the places in the proof where the sum-product theorem is applied). This is done partly to highlight a reading of the proof which has recognizable connections with more "classical" analytic number theory.

– We also include a full proof of one of the two basic ingredients from additive combinatorics that occur in the proof of Theorem 1.1. This is a version of the Balog–Szemerédi–Gowers Theorem (see Theorem 2.2 below), for which Schoen has recently given a short proof (see [10]); our presentation is based on an unpublished note of B. Green. This proof also has a clear probabilistic flavor, and thus fits our presentation very well. (On the other, we only quote the sum-product theorem over finite fields of Bourgain, Katz and Tao [1], which is the other key ingredient from additive combinatorics.)

– On a more technical level, we use the same basic probabilistic lemma to verify the assumptions in the two applications of the Balog–Szemerédi–Gowers Theorem in the proof (see Section 4), and we streamline or uniformize a few other small steps. This should hopefully make the ideas easier to memorize or digest.

Notation

We use $f = \mathrm{O}(g)$ and $f \ll g$ (or $g \gg f$) synonymously: for functions f and g defined on a set X, this means that there exists a real number $\mathrm{C} \geqslant 0$, called sometimes the *implied constant*, such that $|f(x)| \leqslant g(x)$ for all $x \in \mathrm{X}$.

We denote by $|\mathrm{X}|$ the cardinality of a set X.

We denote by $\mathbf{1}_{\mathrm{Y}}$ the characteristic function of a subset Y of a set X.

We note that although we did not attempt to keep track of the constants in the final estimate, we have done so for the "easier" steps. The values of these constants (e.g. in Proposition 6.1) are of course not very important in themselves.

Acknowledgements

We thank B. Green for sending his account of Schoen's result. We also especially thank all the students of the "Additive Combinatorics" class for their interest and active participation in the course, and C. Bortolotto for organizing the exercise sessions. Thanks to A. Gamburd for sending the link to Bourgain's talk [2] and to I. Shkredov for pointing out his improved bound in [11].

2. Preliminaries

We summarize here the background results used in the proof of Theorem 1.1. This section can be skipped until needed during the proof of the theorem.

Lemma 2.1. *Let* X *be a bounded non-negative random variable. Let* $\mathrm{M} \geqslant 0$ *be such that* $\mathrm{X} \leqslant \mathrm{M}$*. Assume that*

$$\mathbf{E}(\mathrm{X}) \geqslant (1-\delta)\mathrm{M}$$

for some $\delta > 0$*. We then have*

$$\mathbf{P}\Big(\mathrm{X} \geqslant (1-\gamma)\mathrm{M}\Big) \geqslant 1 - \frac{\delta}{\gamma}$$

for any γ *such that* $0 < \gamma \leqslant 1$*.*

In particular, if $\mathbf{E}(\mathrm{X}) \geqslant \alpha^{-1}\mathrm{M}$ *for some* $\alpha \geqslant 1$*, then*

$$\mathbf{P}\Big(\mathrm{X} \geqslant \frac{\mathrm{M}}{2\alpha}\Big) \geqslant \frac{1}{2\alpha}. \tag{2.1}$$

Proof. We use Chebychev's inequality to obtain the complementary upper-bound:

$$\mathbf{P}\Big(\mathrm{X} \leqslant (1-\gamma)\mathrm{M}\Big) = \mathbf{P}(\mathrm{M} - \mathrm{X} \geqslant \gamma\mathrm{M}) \leqslant \frac{\mathbf{E}(\mathrm{M}-\mathrm{X})}{\gamma\mathrm{M}} \leqslant \frac{\delta}{\gamma}.$$

In the final assertion, we have $1-\delta = \alpha^{-1}$ and $1-\gamma = 1 - \frac{1}{2}\alpha^{-1}$, so that

$$1 - \frac{\delta}{\gamma} = \frac{\frac{1}{2}\alpha^{-1}}{1-\frac{1}{2}\alpha^{-1}} \geqslant \frac{1}{2\alpha},$$

and the second inequality follows. □

We now discuss the version of the Balog–Szemerédi–Gowers Theorem that we will use. We first fix some notation, to be used throughout.

Given a group G (not necessarily abelian, although this will be the case in the applications below) and finite subsets A and $\mathrm{B} \subset \mathrm{G}$, we denote by $r_{\mathrm{A}\cdot\mathrm{B}}$ the *representation function* for the product set $\mathrm{A}\cdot\mathrm{B} = \{ab \mid (a,b) \in \mathrm{A}\times\mathrm{B}\}$, namely

$$r_{\mathrm{A}\cdot\mathrm{B}}(x) = \sum_{\substack{(a,b)\in\mathrm{A}\times\mathrm{B}\\ ab=x}} 1.$$

This function satisfies $0 \leqslant r_{\mathrm{A}\cdot\mathrm{B}}(x) \leqslant |\mathrm{A}|$ for all $x \in \mathrm{G}$, and

$$\sum_{x\in\mathrm{G}} r_{\mathrm{A}\cdot\mathrm{B}}(x) = |\mathrm{A}||\mathrm{B}|.$$

Moreover, its second moment is the so-called *multiplicative energy* (or just *energy*) of (A,B), which we denote $\mathrm{E}(\mathrm{A},\mathrm{B})$:

$$\mathrm{E}(\mathrm{A},\mathrm{B}) = \sum_{x\in\mathrm{G}} r_{\mathrm{A}\cdot\mathrm{B}}(x)^2 = |\{(a_1,a_2,b_1,b_2) \in \mathrm{A}^2\times\mathrm{B}^2 \mid a_1b_1 = a_2b_2\}|.$$

If A and B are non-empty, we denote by $e(\mathrm{A},\mathrm{B})$ the normalized energy, defined by

$$e(\mathrm{A},\mathrm{B}) = \frac{\mathrm{E}(\mathrm{A},\mathrm{B})}{(|\mathrm{A}||\mathrm{B}|)^{3/2}}.$$

Finally, we denote by A^{-1} the set of inverses of elements of A. If G is abelian, then since $ab = cd$ is equivalent to $ac^{-1} = db^{-1}$, it follows that $\mathrm{E}(\mathrm{A},\mathrm{A}) = \mathrm{E}(\mathrm{A},\mathrm{A}^{-1})$.

Theorem 2.2. *Let* G *be a group and* $\mathrm{A} \subset \mathrm{G}$ *a non-empty finite subset. Let* $\alpha \geqslant 1$ *be such that* $e(\mathrm{A}) \geqslant \alpha^{-1}$*. There exists a subset* $\mathrm{B} \subset \mathrm{A}$ *such that*

$$|\mathrm{B}| \geqslant \frac{|\mathrm{A}|}{4\alpha}, \qquad |\mathrm{B} \cdot \mathrm{B}^{-1}| \leqslant 2^{14}\alpha^6|\mathrm{B}|, \tag{2.2}$$

where the implied constant is absolute.

We will give the proof below.

The last (and crucial) part of the proof is the sum-product theorem of Bourgain, Katz and Tao [1].

Theorem 2.3 (Bourgain–Katz–Tao)**.** *For any* $\gamma > 0$*, there exists* $\delta > 0$ *such that for any prime number* p *and any set* $\mathrm{A} \subset \mathbf{F}_p$ *such that* $|\mathrm{A}| \leqslant p^{1-\gamma}$*, we have*

$$\max(|\mathrm{A}+\mathrm{A}|, |\mathrm{A}\cdot\mathrm{A}|) \gg |\mathrm{A}|^{1+\delta}, \tag{2.3}$$

where the implied constant depends only on γ*.*

Remark 2.4. The original version of the theorem includes also the assumption that $|\mathrm{A}| \geqslant p^{\gamma}$, but this was found to be unnecessary by Konyagin (although it would pose no problem in the application to Theorem 1.1). Two proofs, written in similar style to this paper, can be found in the lecture notes [8, § 4.2] (besides the proof in [1], these notes contain a proof based on ideas of Breuillard [6] related to growth in the affine-linear group).

We finish this section by giving the proof of Theorem 2.2, following essentially a write-up by B. Green of the argument of Schoen [10]. Again, readers who want to focus on the proof of Theorem 1.1 may skip to the beginning of the next section.

The key step is to find a large subset X of A such that the elements of $\mathrm{X}\cdot\mathrm{X}^{-1}$ have a large number of representations as elements of $\mathrm{A}\cdot\mathrm{A}^{-1}$. The precise statement is the following:

Proposition 2.5. *Let* G *be a group and* $\mathrm{A} \subset \mathrm{G}$ *a non-empty finite subset. Let* $\alpha \geqslant 1$ *be such that* $e(\mathrm{A}) \geqslant \alpha^{-1}$*. Fix a real number* δ *such that* $0 < \delta < 1$*. Denote by* r *the representation function for* $\mathrm{A}\cdot\mathrm{A}^{-1}$*.*

There exists $x \in \mathrm{G}$ *such that*

$$|\mathrm{A} \cap \mathrm{A}\cdot x| \geqslant \frac{|\mathrm{A}|}{2\alpha} \tag{2.4}$$

and

$$\left|\left\{(a,b) \in (\mathrm{A}\cap\mathrm{A}\cdot x)^2 \mid r(ab^{-1}) \geqslant \frac{\delta|\mathrm{A}|}{2\alpha^2}\right\}\right| \geqslant (1-\delta)|\mathrm{A}\cap\mathrm{A}\cdot x|^2. \tag{2.5}$$

Proof. The key idea is to take x "at random", but not according to the uniform probability measure on G. Rather, we pick a given element x with probability proportional to $r(x)$. More precisely, since

$$\sum_{x\in\mathrm{G}} r(x) = |\mathrm{A}||\mathrm{A}^{-1}| = |\mathrm{A}|^2,$$

we let X be a G-valued random variable such that

$$\mathbf{P}(\mathrm{X} = x) = \frac{r(x)}{|\mathrm{A}|^2}$$

for any $x \in \mathrm{G}$. We further denote $\mathrm{B} = \mathrm{A} \cap \mathrm{A} \cdot \mathrm{X}$, which is a random subset of G, contained in A.

Let $\gamma > 0$ be a parameter to be chosen later. We define

$$\mathrm{Y} = \{(a,b) \in \mathrm{A} \times \mathrm{A} \mid r(ab^{-1}) < \gamma |\mathrm{A}|\}.$$

We will show that for $\gamma = \delta/(2\alpha^2)$, the inequality

$$\mathbf{E}\Big(|\mathrm{B}|^2 - \delta^{-1}|(\mathrm{B}\times\mathrm{B}) \cap \mathrm{Y}|\Big) \geqslant \frac{|\mathrm{A}|^2}{2\alpha^2} \tag{2.6}$$

holds. It implies the existence of some element $x \in \mathrm{G}$ such that

$$|\mathrm{A} \cap \mathrm{A}\cdot x|^2 - \delta^{-1}|(\mathrm{A}\cap \mathrm{A}\cdot x)^2 \cap \mathrm{Y}| \geqslant \frac{|\mathrm{A}|^2}{2\alpha^2},$$

and from this we deduce, on the one hand, that $|\mathrm{A}\cap\mathrm{A}\cdot x|^2 \geqslant |\mathrm{A}|^2/(2\alpha^2)$, which implies (2.4), and on the other hand that

$$|(\mathrm{A}\cap \mathrm{A}\cdot x)^2 \cap \mathrm{Y}| \leqslant \delta |\mathrm{A}\cap \mathrm{A}\cdot x|^2,$$

which is equivalent to (2.5).

To prove (2.6), we first find a lower-bound for $\mathbf{E}(|\mathrm{B}|^2)$. By the Cauchy–Schwarz inequality, we have $\mathbf{E}(|\mathrm{B}|^2) \geqslant \mathbf{E}(|\mathrm{B}|)^2$, and the expectation of the size of B is

$$\mathbf{E}(|\mathrm{B}|) = \sum_{a\in\mathrm{A}} \mathbf{P}(a \in \mathrm{A}\cdot\mathrm{X}) = \sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{A}} \mathbf{P}(\mathrm{X} = b^{-1}a) = \frac{1}{|\mathrm{A}|^2}\sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{A}} r(b^{-1}a),$$

by definition of X. By replacing $r(b^{-1}a)$ by its definition, we compute

$$\frac{1}{|\mathrm{A}|^2}\sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{A}} r(b^{-1}a) = \frac{1}{|\mathrm{A}|^2}\sum_{a\in\mathrm{A}}\sum_{b\in\mathrm{A}} \sum_{\substack{(x,y)\in\mathrm{A}^2 \\ xy^{-1}=b^{-1}a}} 1 = \frac{\mathrm{E}(\mathrm{A},\mathrm{A})}{|\mathrm{A}|^2} = |\mathrm{A}|e(\mathrm{A}).$$

Using the assumption $e(\mathrm{A}) \geqslant \alpha^{-1}$, we therefore get the lower bound

$$\mathbf{E}(|\mathrm{B}|^2) \geqslant \frac{|\mathrm{A}|^2}{\alpha^2}.$$

We now handle separately an upper bound for the expectation of $(\mathrm{B}\times\mathrm{B})\cap\mathrm{Y}$. We simply write

$$\mathbf{E}(|(\mathrm{B}\times\mathrm{B})\cap\mathrm{Y}|) \leqslant |\mathrm{A}|^2 \max_{(a,b)\in\mathrm{Y}} \mathbf{P}(\{a,b\}\subset \mathrm{B}),$$

and estimate the probability that $\{a,b\} \subset \mathrm{B}$ for each $(a,b) \in \mathrm{Y}$ separately. Since $\mathrm{Y} \subset \mathrm{A}^2$, this is

$$\mathbf{P}(a \in \mathrm{B} \text{ and } b \in \mathrm{B}) = \mathbf{P}(a \in \mathrm{A}\cdot\mathrm{X} \text{ and } b \in \mathrm{A}\cdot\mathrm{X}) = \mathbf{P}(\mathrm{X} \in \mathrm{A}^{-1}\cdot a \cap \mathrm{A}^{-1}\cdot b).$$

From the crude bound $r(x) \leqslant |\mathrm{A}|$, it follows that $\mathbf{P}(\mathrm{X} = x) \leqslant 1/|\mathrm{A}|$ for any $x \in \mathrm{G}$, and we deduce that

$$\mathbf{P}(\mathrm{X} \in \mathrm{A}^{-1}\cdot a \cap \mathrm{A}^{-1}\cdot b) \leqslant \frac{1}{|\mathrm{A}|}|\mathrm{A}^{-1}\cdot a \cap \mathrm{A}^{-1}\cdot b|.$$

We now note that

$$|\mathrm{A}^{-1}\cdot a \cap \mathrm{A}^{-1}\cdot b| = |\{(x,y) \in \mathrm{A}^2 \mid xy^{-1} = ab^{-1}\}|$$

(because of the bijection f which sends an element $w \in \mathrm{A}^{-1} \cdot a \cap \mathrm{A}^{-1} \cdot b$ to (aw^{-1}, bw^{-1}), with inverse $(x, y) \mapsto a^{-1}x = b^{-1}y$). Thus we get

$$\mathbf{P}(a \in \mathrm{B} \text{ and } b \in \mathrm{B}) \leqslant \frac{1}{|\mathrm{A}|} \sum_{\substack{(x,y)\in \mathrm{A}^2 \\ xy^{-1}=ab^{-1}}} 1 = \frac{r(ab^{-1})}{|\mathrm{A}|},$$

and by definition of Y, this is $< \gamma|\mathrm{A}|$. Thus we have

$$\mathbf{E}\Big(|\mathrm{B}|^2 - \delta^{-1}|(\mathrm{B}\times\mathrm{B}) \cap \mathrm{Y}|\Big) \geqslant \frac{|\mathrm{A}|^2}{\alpha^2} - \frac{\gamma|\mathrm{A}|^2}{\delta},$$

and this is $\geqslant |\mathrm{A}|^2/(2\alpha^2)$ if we take $\gamma = \delta/(2\alpha^2)$, as claimed. □

Proof of Theorem 2.2. We apply Proposition 2.5 with $\delta = 1/10$; we denote by C the set $\mathrm{A} \cap \mathrm{A} \cdot x$ which it provides, and let

$$\mathrm{Y} = \Big\{ y \in \mathrm{G} \mid r(y) \geqslant \frac{\delta|\mathrm{A}|}{2\alpha^2} \Big\},$$

where r is again the representation function for $\mathrm{A} \cdot \mathrm{A}^{-1}$. We note that

$$|\mathrm{Y}| \leqslant 20\alpha^2|\mathrm{A}| \tag{2.7}$$

by Chebychev's inequality. Further, for any element $a \in \mathrm{A}$, we denote by $\mathrm{N}(a)$ the set of $b \in \mathrm{C}$ such that $ab^{-1} \in \mathrm{Y}$.

We have $0 \leqslant |\mathrm{N}(c)| \leqslant |\mathrm{C}|$ for any $c \in \mathrm{C}$; moreover, by (2.5), we have

$$\frac{1}{|\mathrm{C}|}\sum_{c\in\mathrm{C}} |\mathrm{N}(c)| \geqslant (1-\delta)|\mathrm{C}|,$$

and this implies that $\mathrm{N}(c)$ must often be quite close to its maximal value. Precisely, from Lemma 2.1 (with X the random variable $c \mapsto \mathrm{N}(c)$ on C with uniform probability), we get

$$|\{c \in \mathrm{C} \mid \mathrm{N}(c) \geqslant (1-\gamma)|\mathrm{C}|\}| \geqslant \Big(1 - \frac{\delta}{\gamma}\Big)|\mathrm{C}|,$$

whenever $0 < \gamma < 1$. Taking $\gamma = \sqrt{\delta}$, we find that there are at least $(1-\sqrt{\delta})|\mathrm{C}|$ elements of C such that $|\mathrm{N}(c)| \geqslant (1-\sqrt{\delta})|\mathrm{C}|$.

Let B be the subset of C (hence of A) defined by this condition on $\mathrm{N}(c)$; since Proposition 2.5 implies that $|\mathrm{C}| \geqslant |\mathrm{A}|/(2\alpha)$, we already get

$$|\mathrm{B}| \geqslant (1-\sqrt{\delta})|\mathrm{C}| \geqslant \frac{|\mathrm{C}|}{2} \geqslant \frac{|\mathrm{A}|}{4\alpha}.$$

To conclude the proof, we claim that

$$\mathrm{B} \cdot \mathrm{B}^{-1} \subset \Big\{ x \in \mathrm{G} \mid s(x) \geqslant \frac{|\mathrm{C}|}{3} \Big\}, \tag{2.8}$$

where s is the representation function for $\mathrm{Y} \cdot \mathrm{Y}^{-1}$. Assuming this, we observe that the right-hand set satisfies

$$\Big|\Big\{ x \in \mathrm{G} \mid s(x) \geqslant \frac{|\mathrm{C}|}{3} \Big\}\Big| \leqslant \frac{3|\mathrm{Y}|^2}{|\mathrm{C}|}$$

(by Chebychev's inequality again). Using $|\mathrm{C}| \geqslant |\mathrm{A}|/(2\alpha)$ together with (2.7), we deduce

$$|\mathrm{B}\cdot\mathrm{B}^{-1}| \leqslant \frac{3|\mathrm{Y}|^2}{|\mathrm{C}|} \leqslant 6\cdot 20^2\cdot\alpha^5|\mathrm{A}| \leqslant 4\cdot 6\cdot 20^2\cdot\alpha^6|\mathrm{B}| \leqslant 2^{14}|\mathrm{B}|,$$

which finishes the proof of the theorem.

To prove (2.8), pick any a and b in B; we need a lower bound for $s(ab^{-1})$, or in other words for the size of the set

$$\{(u,v)\in \mathrm{Y}\times\mathrm{Y} \mid uv^{-1}=ab^{-1}\}.$$

There is an injective map

$$\mathrm{N}(a)\cap\mathrm{N}(b) \to \{(u,v)\in \mathrm{Y}\times\mathrm{Y} \mid uv^{-1}=ab^{-1}\}$$

defined by $f(z)=(az^{-1},bz^{-1})$ (the crucial point here is that this map is well-defined: we have $(az^{-1},bz^{-1})\in \mathrm{Y}\times\mathrm{Y}$ by definition of $\mathrm{N}(a)$ and $\mathrm{N}(b)$). Hence $s(ab^{-1})\geqslant|\mathrm{N}(a)\cap\mathrm{N}(b)|$. But, by definition, $|\mathrm{N}(a)|$ and $|\mathrm{N}(b)|$ are very large, and so is their intersection. In fact, we get

$$|\mathrm{N}(a)\cap\mathrm{N}(b)| \geqslant (1-2\sqrt{\delta})|\mathrm{C}| \geqslant \frac{|\mathrm{C}|}{3},$$

(recall that $\delta=1/10$), so that $s(ab^{-1})\geqslant|\mathrm{C}|/3$, as desired. □

3. Two probabilistic constructions

We already mentioned that we will present the proof of Theorem 1.1 in probabilistic language. This relies on two elementary constructions which we present here, in greater generality than required.

We consider a finite group G. Given a G-valued random variable X (defined on some probability space Ω which we need not specify precisely), we will denote by ϱ_{X} its "density" function, i.e., $\varrho_{\mathrm{X}}\colon \mathrm{G}\to\mathbf{R}$ is the function such that $\varrho_{\mathrm{X}}(x)=\mathbf{P}(\mathrm{X}=x)$ for all $x\in\mathrm{X}$.

Stepping. We say that a G-valued random variable Y is a *stepping* of X if $\mathrm{Y}=\mathrm{X}_1\mathrm{X}_2^{-1}$, where $(\mathrm{X}_1,\mathrm{X}_2)$ are independent random variables, both independent of X and distributed like X. In particular, X and Y are then independent. We have

$$\varrho_{\mathrm{Y}}(y)=\mathbf{P}(\mathrm{Y}=y)=\mathbf{P}(\mathrm{X}_1\mathrm{X}_2^{-1}=y)=\sum_{x\in\mathrm{G}}\mathbf{P}(\mathrm{X}=x)\mathbf{P}(\mathrm{X}=x^{-1}y),$$

and in particular

$$\varrho_{\mathrm{Y}}(0)=\sum_{x\in\mathrm{G}}\mathbf{P}(\mathrm{X}=x)^2. \tag{3.1}$$

Applying the Cauchy–Schwarz inequality to the formula for $\varrho_{\mathrm{Y}}(x)$, we see that $\varrho_{\mathrm{Y}}(x)\leqslant\varrho_{\mathrm{Y}}(0)$ for all $x\in\mathrm{G}$.

Remark 3.1. In additive notation, we have $\mathrm{Y}=\mathrm{X}_1-\mathrm{X}_2$ with $(\mathrm{X},\mathrm{X}_1,\mathrm{X}_2)$ independent and identically distributed.

Peaking. We now assume that G is commutative, with additive notation, and we denote by $\widehat{\mathrm{G}}$ its character group. For any G-valued random variable X, we denote by φ_{X} the "characteristic function" of X (in the probabilistic sense, hence essentially its Fourier transform), namely the function on $\widehat{\mathrm{G}}$ defined by

$$\varphi_{\mathrm{X}}(\chi) = \mathbf{E}(\chi(\mathrm{X}))$$

for $\chi \in \widehat{\mathrm{G}}$. We have $\varphi_{-\mathrm{X}} = \overline{\varphi_{\mathrm{X}}}$, and if X_1 and X_2 are independent, then $\varphi_{\mathrm{X}_1+\mathrm{X}_2} = \varphi_{\mathrm{X}_1}\varphi_{\mathrm{X}_2}$.

Let now $\mathrm{Y} = \mathrm{X}_1 - \mathrm{X}_2$ be a stepping of X. According to the above, we have $\varphi_{\mathrm{Y}} = |\varphi_{\mathrm{X}}|^2$. In particular, since $\varphi_{\mathrm{Y}} = |\varphi_{\mathrm{X}}|^2 \geqslant 0$, and since $\varphi_{\mathrm{Y}}(0) = 1$, we can consider a random variable $\widehat{\mathrm{Y}}$ on $\widehat{\mathrm{G}}$ such that

$$\mathbf{P}(\widehat{\mathrm{Y}} = \chi) = \frac{\varphi_{\mathrm{Y}}(\chi)}{\mathrm{M}_{\mathrm{X}}} = \frac{|\varphi_{\mathrm{X}}(\chi)|^2}{\mathrm{M}_{\mathrm{X}}}$$

for $\chi \in \widehat{\mathrm{G}}$, where

$$\mathrm{M}_{\mathrm{X}} = \sum_{\chi \in \widehat{\mathrm{G}}} |\varphi_{\mathrm{X}}(\chi)|^2.$$

Moreover, we may (and do) insist that $\widehat{\mathrm{Y}}$ is independent from $(\mathrm{X}, \mathrm{X}_1, \mathrm{X}_2)$, hence also from Y. (Similarly, whenever we consider $\widehat{\mathrm{Z}}$ for some other random variable Z, it will be understood that $\widehat{\mathrm{Z}}$ is independent of any previously described random variables.)

Intuitively, the random variable $\widehat{\mathrm{Y}}$ emphasizes the characters χ where $\varphi_{\mathrm{X}}(\chi)$ is large, and for this reason we will say that $\widehat{\mathrm{Y}}$ is a *peaking* of Y, or of X.

Remark 3.2. If $\mathrm{G} = \mathbf{Z}/q\mathbf{Z}$ for some integer $q \geqslant 1$, we can identify as usual the character group with G by associating to $a \in \mathbf{Z}/q\mathbf{Z}$ the character $x \mapsto e(ax/q)$. Thus we also identify the characteristic function φ_{X} with a function $\mathbf{Z}/q\mathbf{Z} \to \mathbf{C}$, with

$$\varphi_{\mathrm{X}}(a) = \mathbf{E}\Big(e\Big(\frac{a\mathrm{X}}{q}\Big)\Big).$$

Steppings and peakings are related by a simple but crucial formula, which reflects the Fourier duality. We identify as usual the dual group of $\widehat{\mathrm{G}}$ with G, the element $x \in \mathrm{G}$ corresponding to the character $\chi \mapsto \chi(x)$ of $\widehat{\mathrm{G}}$.

Lemma 3.3. *Let* G *be a finite commutative group. For any* G*-valued random variable* X*, with stepping* Y *and peaking* $\widehat{\mathrm{Y}}$*, and for any* $y \in \mathrm{G}$*, we have*

$$\varrho_{\mathrm{Y}}(y) = \frac{\mathrm{M}_{\mathrm{X}}}{|\mathrm{G}|}\varphi_{\widehat{\mathrm{Y}}}(y),$$

where the characteristic function of $\widehat{\mathrm{Y}}$ *is identified with a function on* G*.*

Proof. We use the orthogonality of characters to represent the (set-theoretic!) characteristic function of an element $y \in \mathrm{G}$ by

$$\frac{1}{|\mathrm{G}|}\sum_{\chi \in \widehat{\mathrm{G}}} \chi(x-y) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y, \end{cases}$$

and get

$$\varrho_Y(y) = \mathbf{E}\Big(\frac{1}{|G|}\sum_{\chi\in\widehat{G}}\chi(Y-y)\Big) = \frac{1}{|G|}\sum_{\chi\in\widehat{G}}\chi(-y)\varphi_Y(\chi) = \frac{M_X}{|G|}\varphi_{\widehat{Y}}(-y),$$

by definition of $\widehat{Y}$. This proves the lemma since $\varrho_Y(-y) = \varrho_Y(y)$. □

In particular, we note the formula

$$\varrho_Y(0) = \frac{M_X}{|G|}. \tag{3.2}$$

Remark 3.4. If X is uniformly distributed on G, then Y is also uniformly distributed on G, and $\widehat{Y}$ is a Dirac mass at the unit element 1 of G. Conversely, if X is a Dirac mass at some $x \in G$, then Y is a Dirac mass at 1, and $\widehat{Y}$ is uniformly distributed on $\widehat{G}$.

4. Probabilistic lemmas

In order to apply Theorem 2.2, we will use two lemmas giving probabilistic conditions that guarantee large energy. We use the definition of a "stepping" of a random variable from the previous section.

Lemma 4.1. *Let* G *be a finite group and let* A *be a non-empty subset of* G. *Let* X *be a* G*-valued random variable and* Y *a stepping of* X. *We assume that* $\beta \geqslant 1$ *is such that*

$$\mathbf{E}(r_{A\cdot A^{-1}}(X)) \geqslant \beta^{-1}|A|.$$

We then have

$$e(A) \geqslant \frac{1}{4\beta^4\varrho_Y(0)|A|}.$$

Proof. Let

$$L = \{x \in G \mid r_{A\cdot A^{-1}}(x) \geqslant \tfrac{1}{2}\beta^{-1}|A|\},$$

so that we have the lower-bound

$$E(A) = \sum_{x\in G} r_{A\cdot A^{-1}}(x)^2 \geqslant \sum_{x\in L} r_{A\cdot A^{-1}}(x)^2 \geqslant \beta^{-2}|A|^2|L|.$$

Noting that $r_{A\cdot A^{-1}}(x) \leqslant |A|$ for all x, the assumption implies that

$$\mathbf{P}(L) = \mathbf{P}\Big(r_{A\cdot A^{-1}}(X) \geqslant \frac{|A|}{2\beta}\Big) \geqslant \frac{1}{2\beta}$$

(see (2.1)), but the Cauchy–Schwarz inequality and positivity imply that

$$\mathbf{P}(L) = \sum_{x\in L}\mathbf{P}(X=x) \leqslant |L|^{1/2}\Big(\sum_{x\in G}\mathbf{P}(X=x)^2\Big)^{1/2} = |L|^{1/2}\varrho_Y(0)^{1/2},$$

and hence $|L| \geqslant (2\beta)^{-2}\varrho_Y(0)^{-1}$. The previous lower-bound gives

$$E(A) \geqslant 2^{-2}\beta^{-4}\varrho_Y(0)^{-1}|A|^2,$$

which implies the desired result. □

The second and final lemma uses this to conclude that the energy of the set of "elements with large probability" will be big if those sets are of "typical" size.

Lemma 4.2. *Let* G *be a finite group. Let* X *be a* G*-valued random variable and let* $\mathrm{Y} = \mathrm{X}_1\mathrm{X}_2^{-1}$ *be a stepping of* X*. Let* $\alpha \geqslant 1$ *and define*

$$\mathrm{A} = \Big\{ x \in \mathrm{G} \mid \mathbf{P}(\mathrm{Y} = x) \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{\alpha} \Big\}.$$

Let $\mathrm{B} \subset \mathrm{A}$ *and let* $\beta > 0$ *be such that*

$$|\mathrm{B}| \geqslant \frac{1}{\beta \varrho_{\mathrm{Y}}(0)}.$$

We have then

$$e(\mathrm{B}) \geqslant \frac{1}{4\alpha^9\beta^4}.$$

Proof. Let $r = r_{\mathrm{B}\cdot\mathrm{B}^{-1}}$ be the representation function for $\mathrm{B}\cdot\mathrm{B}^{-1}$. We have

$$\mathbf{E}(r(\mathrm{Y})) = \sum_{a,b\in\mathrm{B}} \mathbf{P}(\mathrm{Y} = ab^{-1}) = \sum_{a,b\in\mathrm{A}} \mathbf{P}(\mathrm{X}_1 a^{-1} = \mathrm{X}_2 b^{-1}),$$

and this implies that

$$\begin{aligned}\mathbf{E}(r(\mathrm{Y})) &= \sum_{y\in\mathrm{G}} \sum_{a,b\in\mathrm{B}} \mathbf{P}(\mathrm{X}_1 a^{-1} = y \text{ and } \mathrm{X}_2 b^{-1} = y) \\ &= \sum_{y\in\mathrm{G}} \sum_{a,b\in\mathrm{B}} \mathbf{P}(\mathrm{X}_1 a^{-1} = y)\mathbf{P}(\mathrm{X}_2 b^{-1} = y) = \sum_{y\in\mathrm{G}} \mathbf{P}(\mathrm{X}_1 \in y\mathrm{B})^2.\end{aligned}$$

The "reversed" Cauchy–Schwarz inequality now shows that for any choice of $f(y) \geqslant 0$ for $y \in \mathrm{G}$, not all zero, we have

$$\mathbf{E}(r(\mathrm{Y})) \geqslant \frac{\mathrm{V}^2}{\mathrm{W}}$$

with

$$\mathrm{V} = \sum_{y\in\mathrm{G}} f(y)\mathbf{P}(\mathrm{X}_1 \in y\mathrm{B}), \qquad \mathrm{W} = \sum_{y\in\mathrm{G}} f(y)^2.$$

We pick $f(y) = \mathbf{P}(\mathrm{X}_2 = y)$; in this case, we have

$$\mathrm{V} = \mathbf{P}(\mathrm{Y} \in \mathrm{B}), \qquad \mathrm{W} = \mathbf{P}(\mathrm{Y} = 0),$$

and therefore

$$\mathbf{E}(r(\mathrm{Y})) \geqslant \frac{\mathbf{P}(\mathrm{Y} \in \mathrm{B})^2}{\varrho_{\mathrm{Y}}(0)} \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{\alpha^2}|\mathrm{B}|^2,$$

where the last step follows from the assumption that $\mathrm{B} \subset \mathrm{A}$, so that $\mathbf{P}(\mathrm{Y} = y) \geqslant \alpha^{-1}\varrho_{\mathrm{Y}}(0)$ for $y \in \mathrm{B}$. Since we also assumed that $\varrho_{\mathrm{Y}}(0)|\mathrm{B}| \geqslant \beta^{-1}$, this gives $\mathbf{E}(r(\mathrm{Y})) \geqslant \alpha^{-2}\beta^{-1}|\mathrm{B}|$.

Applying Lemma 4.1 to the random variable Y and the set B, we get

$$e(\mathrm{B}) \geqslant \frac{1}{4\alpha^8\beta^4\varrho_{\mathrm{Z}}(0)|\mathrm{B}|},$$

where Z is a stepping of Y. But we have

$$\varrho_{\mathrm{Z}}(0) = \mathbf{P}(\mathrm{Z}=0) = \sum_{y\in \mathrm{G}} \mathbf{P}(\mathrm{Y}=y)^2 \leqslant \mathbf{P}(\mathrm{Y}=0)\sum_{y\in \mathrm{G}} \mathbf{P}(\mathrm{Y}=y) = \mathbf{P}(\mathrm{Y}=0) = \varrho_{\mathrm{Y}}(0),$$

and thus $\varrho_{\mathrm{Z}}(0)|\mathrm{B}| \leqslant \varrho_{\mathrm{Y}}(0)|\mathrm{A}|$, which is $\leqslant \alpha$ by Chebychev's inequality, so we get finally the lower bound

$$e(\mathrm{B}) \geqslant \frac{1}{4\alpha^8\beta^4\varrho_{\mathrm{Y}}(0)|\mathrm{B}|} \geqslant \frac{1}{4\alpha^9\beta^4},$$

as claimed. □

5. Main steps of the proof

We will describe in this section the strategy of the proof of Theorem 1.1, extracting two intermediate steps before the final conclusion.

Step 1. The first step is an estimate for a specific average of values of the discrete Fourier transform of random variables on $\mathbf{F}_p$, which involves the "peaking" of Section 3.

Proposition 5.1. *Let p be a prime number. Let X be an $\mathbf{F}_p$-valued random variable, and let $\mathrm{Y} = \mathrm{X}_1 - \mathrm{X}_2$ be a stepping of X and $\widehat{\mathrm{Y}}$ a peaking of X.*

Let $\eta > 0$ be a real number. There exists $\beta > 0$, depending only on η, such that

$$\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) \ll \varrho_{\mathrm{X}}(0) + \varrho_{\mathrm{Y}}(0)^{\beta} + \frac{p^{-1+\eta}}{\varrho_{\mathrm{Y}}(0)}. \tag{5.1}$$

Remark 5.2. (1) To get a feeling for this inequality, note the obvious lower bounds

$$\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) \geqslant \mathbf{P}(\mathrm{X}=0), \qquad \mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) \geqslant \mathbf{P}(\widehat{\mathrm{Y}}=0).$$

The term $\varrho_{\mathrm{X}}(0)$ on the right-hand side of (5.1) accounts for the first of these, and the third term accounts for (a quantity larger than) the second, since by (3.2), we have

$$\mathbf{P}(\widehat{\mathrm{Y}}=0) = \frac{1}{\mathrm{M}_{\mathrm{X}}} = \frac{p^{-1}}{\varrho_{\mathrm{Y}}(0)}.$$

(2) Although the bound (5.1) may look conventional enough, it is in its proof that additive combinatorics is crucial. In other words: if (5.1) could be proved "with classical means", i.e. without invoking the sum-product phenomenon, or the Balog–Szemerédi–Gowers Theorem, or other results from additive combinatorics, then this would give a "classical" proof of Theorem 1.1.

(3) In "concrete" terms, without probabilistic notation, the quantity $\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2)$ is the average

$$\frac{1}{\mathrm{M}_{\mathrm{X}}^2}\sum_{x\in\mathbf{F}_p}\sum_{a\in\mathbf{F}_p} \varrho_{\mathrm{X}}(x)|\varphi_{\mathrm{X}}(a)|^2|\varphi_{\mathrm{X}}(ax)|^2.$$

From an analytic number theory point of view, this can be interpreted as a kind of "amplified" average of the values of $|\varphi_{\mathrm{X}}|^2$. To see why this can be useful, take the random variable X to be uniformly distributed over a subgroup H of $\mathbf{F}_p^\times$. Observe (as we will repeat

later) that $\varphi_{\mathrm{X}}(ah) = \varphi_{\mathrm{X}}(a)$ for any $h \in \mathrm{H}$ and $a \neq 0$; it follows that $\varrho_{\mathrm{Y}}(0) = 1/|\mathrm{H}|$, and a simple computation shows that $\mathrm{X}\widehat{\mathrm{Y}}$ is distributed like Y and that

$$\mathrm{M}_{\mathrm{X}} = \sum_{a \in \mathbf{F}_p} \Big| \frac{1}{p} \sum_{x \in \mathrm{H}} e\Big(\frac{ax}{p}\Big) \Big|^2 = \frac{p}{|\mathrm{H}|}.$$

Therefore, for any $a \in \mathbf{F}_p^\times$, we have a lower bound

$$\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) \geqslant |\varphi_{\mathrm{X}}(a)|^2 \mathbf{P}(\mathrm{X}\widehat{\mathrm{Y}} \in \mathrm{H}) \geqslant |\varphi_{\mathrm{X}}(a)|^2 \times |\mathrm{H}| \frac{|\varphi_{\mathrm{X}}(a)|^2}{\mathrm{M}_{\mathrm{X}}} = |\varphi_{\mathrm{X}}(a)|^4 \frac{|\mathrm{H}|^2}{p}.$$

This shows that even the trivial bound $\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) \leqslant 1$ is sufficient to deduce that $|\varphi_{\mathrm{X}}(a)|^4 \leqslant p|\mathrm{H}|^{-2}$, which is non-trivial as soon as H has size a bit larger than $\sqrt{p}$ – the same range in which a "direct" use of Gauss sums leads to a non-trivial bound.

Furthermore, if we apply Proposition 5.1 instead of the trivial bound, with $\eta = \gamma/2$, say, then we get some $\beta > 0$ such that

$$\frac{|\mathrm{H}|^2}{p} |\varphi_{\mathrm{X}}(a)|^4 \ll \frac{1}{|\mathrm{H}|} + \frac{1}{|\mathrm{H}|^{1+\beta}} + \frac{|\mathrm{H}| p^{\eta}}{p}$$

hence

$$|\varphi_{\mathrm{X}}(a)|^4 \ll p^{1-3\gamma} + p^{1-(2+\beta)\gamma} + p^{-\gamma/2},$$

which proves Theorem 1.1 when $|\mathrm{H}| = p^\gamma$ with $\gamma > \max(\frac{1}{3}, \frac{1}{2+\beta})$, hence also for γ slightly smaller than $1/2$. *This is already a highly non-trivial fact.* A result of that type was first proved by Shparlinski [12] (for $|\mathrm{H}|$ a bit larger than $p^{3/7}$), using estimates of Garcia and Voloch on the number of points on Fermat curves over finite fields, also combined with a fourth moment computation.

Step 2. We now describe for which random variables we will apply Proposition 6.1. Let $\mathrm{H} \subset \mathbf{F}_p^\times$ be a multiplicative subgroup. We fix a random variable S which is uniformly distributed on H (so that $\varrho_{\mathrm{S}}(x) = 0$ unless $x \in \mathrm{H}$, in which case $\varrho_{\mathrm{S}}(x) = 1/|\mathrm{H}|$). We denote by $(\mathrm{S}_k)_{k \geqslant 1}$ a sequence of independent random variables, all independent from S and also uniformly distributed on H.

We will consider the random variables

$$\mathrm{X}_k = \mathrm{S}_1 - \mathrm{S}_2 + \cdots + \mathrm{S}_{2k-1} - \mathrm{S}_{2k}$$

for $k \geqslant 1$. Probabilistically, these correspond to a simple random walk on $\mathbf{F}_p$ where the steps are taken alternately from H and from $-\mathrm{H}$ (so the picture could be simplified a bit in the case where $-1 \in \mathrm{H}$, since then each S_i would be distributed in the same way as $-\mathrm{S}_i$, and we would have a "standard" random walk). Note that

$$\varphi_{\mathrm{X}_k}(a) = |\varphi_{\mathrm{S}}(a)|^{2k},$$

by independence; moreover, note that

$$\mathrm{X}_{2k} = (\mathrm{S}_1 - \mathrm{S}_2 + \cdots + \mathrm{S}_{2k-1} - \mathrm{S}_{2k}) - (\mathrm{S}_{2k+2} - \mathrm{S}_{2k+1} + \cdots + \mathrm{S}_{4k} - \mathrm{S}_{4k-1}),$$

which shows that X_{2k} is a stepping of X_k.

For $\nu > 0$, we define the set

$$\Lambda_\nu = \{a \in \mathbf{F}_p \mid |\varphi_{\mathrm{S}}(a)| > p^{-\nu}\}.$$

Note that $0 \in \Lambda_\nu$ in all cases, and that, since

$$\varphi_{\mathrm{S}}(a) = \frac{1}{|\mathrm{H}|}\sum_{x\in \mathrm{H}} e\Big(\frac{ax}{p}\Big),$$

we can restate Theorem 1.1 as claiming the existence of some $\nu > 0$ such that Λ_ν *only* contains 0. This is therefore our objective. The following simple lemma encapsulates the specific property of the distribution of the random variable S.

Lemma 5.3. *For any $x \in \mathrm{H}$, the random variable $x\mathrm{S}$ is uniformly distributed on* H.

In particular the following properties hold:

(1) *For any $a \in \mathbf{F}_p$, we have $\varphi_{\mathrm{S}}(ax) = \varphi_{\mathrm{S}}(a)$, and hence also $\varphi_{\mathrm{X}_k}(ax) = \varphi_{\mathrm{X}_k}(a)$.*
(2) *The set $\Lambda_\nu - \{0\}$ is either empty or is a union of* H*-cosets. In the second case, we have $|\Lambda_\nu| \geqslant |\mathrm{H}|$.*

Proof. The first statement simply reflects the fact that H is a multiplicative subgroup of $\mathbf{F}_p^\times$. The equality $\varphi_{\mathrm{S}}(ax) = \varphi_{\mathrm{S}}(a)$ follows, and it means that $a\mathrm{H} \subset \Lambda_\nu$ whenever $a \in \Lambda_\nu - \{0\}$, which gives the last fact. □

The content of the second step is as follows:

Proposition 5.4. *Let $\theta > 0$ be a real number. If p is a large enough prime number, depending only on θ, then there exist a positive real number $\nu < \frac{1}{2}\theta$, depending only on θ, and an integer $k \geqslant 1$ such that*

$$p^{-1-\theta}|\Lambda_\nu| \leqslant \varrho_{\mathrm{X}_{2k}}(0) \leqslant p^{-1+\theta}|\Lambda_\nu| \tag{5.2}$$

and

$$\mathbf{E}(|\varphi_{\mathrm{X}_k}(\mathrm{X}_k\widehat{\mathrm{X}}_{2k})|^2) \geqslant p^{-10\theta}. \tag{5.3}$$

Step 3. We now conclude the proof of Theorem 1.1. Recall that $|\mathrm{H}| \geqslant p^\gamma$ by assumption; we pick $\theta > 0$ such that $10\theta < \gamma$. Applying Proposition 5.4 and then Proposition 5.1, for some $\eta > 0$ to be determined later, we find random variables $\mathrm{X} = \mathrm{X}_k$ and $\mathrm{Y} = \mathrm{X}_{2k}$ satisfying the bounds (5.2) and such that

$$p^{-10\theta} \leqslant \mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) = \mathbf{E}(|\varphi_{\mathrm{X}_k}(\mathrm{X}_k\widehat{\mathrm{X}}_{2k})|^2) \ll \varrho_{\mathrm{X}}(0) + \varrho_{\mathrm{Y}}(0)^\beta + \frac{p^{-1+\eta}}{\varrho_{\mathrm{Y}}(0)},$$

for some $\beta > 0$.

The first term is easily handled: by induction on k, we find that

$$\mathbf{P}(\mathrm{X}_k = 0) \leqslant \max_{x\in\mathbf{F}_p} \mathbf{P}(\mathrm{S} = x) = \frac{1}{|\mathrm{H}|}$$

for any $k \geqslant 1$, hence the assumption $|\mathrm{H}| \geqslant p^\gamma$ gives

$$p^{-10\theta} \ll p^{-\gamma} + \varrho_{\mathrm{Y}}(0)^\beta + \frac{p^{-1+\eta}}{\varrho_{\mathrm{Y}}(0)}.$$

Using (5.2) to estimate $\varrho_{\mathrm{Y}}(0)$ in terms of $|\Lambda_\nu|$, this becomes

$$p^{-10\theta} \ll p^{-\gamma} + \Big(\frac{|\Lambda_\nu|}{p^{1-\theta}}\Big)^\beta + \frac{p^{\eta+\theta}}{|\Lambda_\nu|}.$$

We always have $|\Lambda_\nu| \leqslant p^{1+2\nu}|\mathrm{H}|^{-1}$ by Chebychev's inequality. Moreover, if we assume that Λ_ν is not reduced to 0, then this set contains at least $|\mathrm{H}| \geqslant p^\gamma$ elements. Recalling that $2\nu < \eta$, we would then get the bounds

$$p^{-10\theta} \ll p^{\beta(2\nu+\theta-\gamma)} + p^{\eta+\theta-\gamma} \ll p^{\beta(\eta+\theta-\gamma)} + p^{\eta+\theta-\gamma},$$

which is impossible for p large enough if η is chosen small enough in terms of γ. Thus we must have $\Lambda_\nu = \{0\}$, and (by definition) this means that

$$\Big|\frac{1}{|\mathrm{H}|}\sum_{x\in\mathrm{H}} e\Big(\frac{ax}{p}\Big)\Big| \leqslant p^{-\nu}$$

for all $a \in \mathbf{F}_p^\times$, provided p is large enough.

6. Completion of the proof

We now prove Propositions 5.1 and 5.4. The sum-product theorem appears decisively in the proof of the first of these, and more precisely in the following key proposition.

Proposition 6.1. *Let p be a prime number. Let X be an $\mathbf{F}_p$-valued random variable, and let $\mathrm{Y} = \mathrm{X}_1 - \mathrm{X}_2$ be a stepping of X as above. Let $\alpha \geqslant 1$ be a real number such that*

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{\alpha}. \tag{6.1}$$

Assuming that

$$\mathbf{P}(\mathrm{X}=0) \leqslant \frac{1}{4\alpha}, \qquad \mathbf{P}(\mathrm{Y}=0) \leqslant \frac{1}{4\alpha}, \tag{6.2}$$

there exists a subset $\mathrm{A} \subset \mathbf{F}_p^\times$ such that

$$\frac{1}{2^{31}\alpha^{10}\varrho_{\mathrm{Y}}(0)} \leqslant |\mathrm{A}| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)}$$

with the property that

$$\max(|\mathrm{A}+\mathrm{A}|, |\mathrm{A}\cdot\mathrm{A}|) \leqslant 2^{878}\alpha^{294}|\mathrm{A}|.$$

Remark 6.2. As already indicated, the constants should really be interpreted as being of the form $c\alpha^d$ for some absolute constants $c > 0$ and $d > 0$.

Remark 6.3. The use of the random variable $\widehat{\mathrm{Y}}$ (which emphasizes values $a \in \mathbf{F}_p$ where $|\varphi_{\mathrm{X}}(a)|^2$ is "large") is reminiscent of the similar use of a non-uniform distribution in the proof of Theorem 2.2.

Proof. We will use frequently the fact that $\varrho_{\mathrm{Y}}(y) \leqslant \varrho_{\mathrm{Y}}(0)$ for all $y \in \mathbf{F}_p$, which we already mentioned.

We define

$$\mathrm{A}_1 = \Big\{ y \in \mathbf{F}_p \mid \varrho_{\mathrm{Y}}(y) \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{8\alpha} \Big\}$$

and $\mathrm{A}_2 = \mathrm{A}_1 - \{0\} \subset \mathbf{F}_p^\times$ (note that $0 \in \mathrm{A}_1$). The main properties of A_2 are given by the next lemma.

Lemma 6.4. *We have*

$$\frac{1}{4\alpha\varrho_{\mathrm{Y}}(0)} \leqslant |\mathrm{A}_2| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)}, \tag{6.3}$$

and the representation function r_2 for $\mathrm{A}_2 \cdot \mathrm{A}_2^{-1}$ satisfies

$$\mathbf{E}(r_2(\mathrm{X})) \geqslant \frac{|\mathrm{A}_2|}{32\alpha^2}. \tag{6.4}$$

Proof. First, simply by Chebychev's inequality, we have

$$|\mathrm{A}_2| \leqslant |\mathrm{A}_1| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)}. \tag{6.5}$$

We now claim that the assumption (6.1), namely

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{\alpha},$$

together with (6.2), implies that

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\, \mathrm{Y}\in \mathrm{A}_1\cap \mathrm{X}^{-1}\mathrm{A}_1}) \geqslant \frac{\varrho_{\mathrm{Y}}(0)}{2\alpha}. \tag{6.6}$$

This is a matter of showing that the contributions to $\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY}))$ from the complementary event, where $\mathrm{X} = 0$ or $\mathrm{Y} \notin \mathrm{A}_1$, or $\mathrm{XY} \notin \mathrm{A}_1$, are small enough. And indeed, first of all the first part of (6.2) gives the upper bound

$$\mathbf{E}\Big(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}=0}\Big) = \varrho_{\mathrm{Y}}(0)\mathbf{P}(\mathrm{X}=0) \leqslant \frac{\varrho_{\mathrm{Y}}(0)}{4\alpha},$$

while

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\ \mathrm{XY}\notin \mathrm{A}_1}) \leqslant \frac{1}{8\alpha}\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) \leqslant \frac{\varrho_{\mathrm{Y}}(0)}{8\alpha}.$$

To bound the last contribution with $\mathrm{X} \neq 0$ and $\mathrm{Y} \notin \mathrm{A}_1$, we write

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\ \mathrm{Y}\notin \mathrm{A}_1}) = \sum_{y\notin \mathrm{A}_1} \mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\, \mathrm{Y}=y}) = \sum_{y\notin \mathrm{A}_1} \mathbf{E}(\varrho_{\mathrm{Y}}(y\mathrm{X})\mathbf{1}_{\mathrm{X}\neq 0,\, \mathrm{Y}=y}).$$

Using the independance of X and Y, we deduce that

$$\begin{aligned}\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\ \mathrm{Y}\notin \mathrm{A}_1}) &= \sum_{y\in \mathbf{F}_p - \mathrm{A}_1} \mathbf{P}(\mathrm{Y}=y)\mathbf{E}(\varrho_{\mathrm{Y}}(y\mathrm{X})\mathbf{1}_{\mathrm{X}\neq 0}) \\ &\leqslant \frac{\varrho_{\mathrm{Y}}(0)}{8\alpha}\mathbf{E}\Big(\sum_{y\notin \mathrm{A}_1} \varrho_{\mathrm{Y}}(y\mathrm{X})\mathbf{1}_{\mathrm{X}\neq 0}\Big) \leqslant \frac{\varrho_{\mathrm{Y}}(0)}{8\alpha}\mathbf{E}\Big(\sum_{y\in \mathbf{F}_p} \varrho_{\mathrm{Y}}(y\mathrm{X})\mathbf{1}_{\mathrm{X}\neq 0}\Big) \leqslant \frac{\varrho_{\mathrm{Y}}(0)}{8\alpha},\end{aligned}$$

using in the last step the fact that, for any given $x \neq 0$, we have

$$\sum_{y\in \mathbf{F}_p} \varrho_{\mathrm{Y}}(yx) = \mathbf{P}(\mathrm{Y}\neq 0) \leqslant 1.$$

We next deduce from (6.6) a lower-bound for $|\mathrm{A}_1|$ complementing the upper-bound (6.5), namely

$$\frac{1}{2\alpha\varrho_{\mathrm{Y}}(0)} \leqslant |\mathrm{A}_1| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)}, \tag{6.7}$$

which in turn implies that $|\mathrm{A}_1| \geqslant 2$ (by (6.2) since $\varrho_{\mathrm{Y}}(0) = \mathbf{P}(\mathrm{Y} = 0)$), and therefore also $|\mathrm{A}_2| = |\mathrm{A}_1| - 1 \geqslant \frac{1}{2}|\mathrm{A}_1|$, hence

$$\frac{1}{4\alpha\varrho_{\mathrm{Y}}(0)} \leqslant |\mathrm{A}_2| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)},$$

Indeed, we obtain (6.7) by noting that, by (6.6), we have

$$\frac{\varrho_{\mathrm{Y}}(0)}{2\alpha} \leqslant \mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\ \mathrm{Y}\in\mathrm{A}_1}) \leqslant \varrho_{\mathrm{Y}}(0)\mathbf{P}(\mathrm{Y}\in\mathrm{A}_1) \leqslant \varrho_{\mathrm{Y}}(0)^2|\mathrm{A}_1|.$$

The next step is to relate the bound (6.6) to the representation function r_2 for $\mathrm{A}_2 \cdot \mathrm{A}_2^{-1}$. For this, we start with the formula

$$\mathbf{E}(r_2(\mathrm{X})) = \sum_{y,z\in\mathrm{A}_2} \mathbf{P}(\mathrm{X} = y^{-1}z) = \sum_{y\in\mathrm{A}_2} \mathbf{E}\Big(\sum_{z\in\mathrm{A}_2} \mathbf{P}(y\mathrm{X} = z)\Big) = \sum_{y\in\mathrm{A}_2} \mathbf{P}(y\mathrm{X}\in\mathrm{A}_2).$$

On the other hand, by independance of X and Y, we have

$$\begin{aligned}\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})\mathbf{1}_{\mathrm{X}\neq 0,\,\mathrm{Y}\in\mathrm{A}_1\cap\mathrm{X}^{-1}\mathrm{A}_1}) &= \sum_{y\in\mathrm{A}_1} \varrho_{\mathrm{Y}}(y)\mathbf{E}(\varrho_{\mathrm{Y}}(y\mathrm{X})\mathbf{1}_{\mathrm{X}\neq 0,\,y\mathrm{X}\in\mathrm{A}_1}) \\ &\leqslant \varrho_{\mathrm{Y}}(0)^2\mathbf{E}\Big(\sum_{y\in\mathrm{A}_1} \mathbf{1}_{\mathrm{X}\neq 0,\ y\mathrm{X}\in\mathrm{A}_1}\Big) = \varrho_{\mathrm{Y}}(0)^2 \sum_{y\in\mathrm{A}_1} \mathbf{P}(\mathrm{X}\neq 0 \text{ and } y\mathrm{X}\in\mathrm{A}_1).\end{aligned}$$

Isolating the contribution of $y = 0 \in \mathrm{A}_1$, we then have

$$\sum_{y\in\mathrm{A}_1} \mathbf{P}(\mathrm{X}\neq 0 \text{ and } y\mathrm{X}\in\mathrm{A}_1) = \mathbf{P}(\mathrm{X}\neq 0) + \mathbf{E}(r_2(\mathrm{X})) \leqslant 1 + \mathbf{E}(r_2(\mathrm{X})),$$

and thus (6.6) implies that

$$\frac{\varrho_{\mathrm{Y}}(0)}{2\alpha} \leqslant \varrho_{\mathrm{Y}}(0)^2\mathbf{E}(r_2(\mathrm{X})) + \varrho_{\mathrm{Y}}(0)^2.$$

The assumption $\mathbf{P}(\mathrm{Y} = 0) = \varrho_{\mathrm{Y}}(0) \leqslant (4\alpha)^{-1}$ (see (6.2)) now leads to the lower-bound

$$\mathbf{E}(r_2(\mathrm{X})) \geqslant \frac{1}{4\alpha\varrho_{\mathrm{Y}}(0)} \geqslant \frac{|\mathrm{A}_2|}{32\alpha^2},$$

concluding the proof. □

Using (6.4), we can apply Lemma 4.1 to the random variable X on $\mathbf{F}_p^\times$, with $\beta = 32\alpha^2$; we obtain

$$e(\mathrm{A}_2) \geqslant \frac{1}{2^{22}\alpha^8\varrho_{\mathrm{Y}}(0)|\mathrm{A}_2|} \geqslant \frac{1}{2^{25}\alpha^9},$$

and therefore, by the Balog–Szemerédi–Gowers Theorem (Theorem 2.2, applied to $\mathrm{A}_2 \subset \mathbf{F}_p^\times$), there exists a subset $\mathrm{A}_3 \subset \mathrm{A}_2$ with

$$|\mathrm{A}_2| \leqslant 4(2^{25}\alpha^9)|\mathrm{A}_3| = 2^{27}\alpha^9|\mathrm{A}_3|, \qquad |\mathrm{A}_3 \cdot \mathrm{A}_3| \leqslant 2^{14}(2^{25}\alpha^9)^6|\mathrm{A}_3| = 2^{164}\alpha^{54}|\mathrm{A}_3|.$$

But we can also control the additive properties of A_3. Precisely, we can apply Lemma 4.2 to the group $\mathbf{F}_p$, the random variables X and Y, and the set $\mathrm{B} = \mathrm{A}_3$, with parameters

$(\alpha,\beta) = (8\alpha, 2^{29}\alpha^{10})$, since $\mathrm{A}_3 \subset \mathrm{A}_1$ and

$$|\mathrm{A}_3| \geqslant \frac{|\mathrm{A}_2|}{2^{27}\alpha^9} \geqslant \frac{1}{2^{29}\alpha^{10}\varrho_{\mathrm{Y}}(0)}$$

thanks to (6.3). The conclusion is that

$$e(\mathrm{A}_3) \geqslant \frac{1}{4(8\alpha)^9(2^{29}\alpha^{10})^4} = \frac{1}{2^{144}\alpha^{49}}.$$

Applying Theorem 2.2 to $\mathrm{A}_3 \subset \mathbf{F}_p$, we find a subset $\mathrm{A}_4 \subset \mathrm{A}_3$ with $|\mathrm{A}_3| \leqslant 4\alpha|\mathrm{A}_4|$ and

$$|\mathrm{A}_4 + \mathrm{A}_4| \leqslant 2^{14}(2^{144}\alpha^{49})^6|\mathrm{A}_4| = 2^{878}\alpha^{294}|\mathrm{A}_4|.$$

Since, in addition, we have

$$|\mathrm{A}_4 \cdot \mathrm{A}_4| \leqslant |\mathrm{A}_3 \cdot \mathrm{A}_3| \leqslant 2^{164}\alpha^{54}|\mathrm{A}_3| \leqslant 2^{166}\alpha^{55}|\mathrm{A}_4|,$$

and

$$\frac{1}{2^{31}\alpha^{10}\varrho_{\mathrm{Y}}(0)} \leqslant \frac{|\mathrm{A}_3|}{4\alpha} \leqslant |\mathrm{A}_4| \leqslant |\mathrm{A}_3| \leqslant \frac{8\alpha}{\varrho_{\mathrm{Y}}(0)},$$

we finally have proved Proposition 6.1 with the set A equal to A_4. □

In order to prove Proposition 5.1, we combine this with a consequence of Lemma 3.3, using Fourier analysis to obtain a "diophantine" interpretation of $\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2)$.

Lemma 6.5. *We have*

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) = \varrho_{\mathrm{Y}}(0)\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2).$$

Proof. Using the formula $\varrho_{\mathrm{Y}}(0) = \mathrm{M}_{\mathrm{X}}/p$ and Lemma 3.3, we have

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) = \varrho_{\mathrm{Y}}(0)\mathbf{E}(\varphi_{\widehat{\mathrm{Y}}}(\mathrm{XY})),$$

and it only remains to appeal to the symmetry formula

$$\mathbf{E}(\varphi_{\widehat{\mathrm{Y}}}(\mathrm{XY})) = \mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2)$$

to conclude. This last identity can be seen as a (very simple) instance of Fubini's formula:

$$\begin{aligned}\mathbf{E}(\varphi_{\widehat{\mathrm{Y}}}(\mathrm{XY})) = \mathbf{E}\Big(\mathbf{E}\Big(e\Big(\frac{\mathrm{XY}\widehat{\mathrm{Y}}}{p}\Big)\Big)\Big) &= \mathbf{E}\Big(\mathbf{E}\Big(e\Big(\frac{\mathrm{X}(\mathrm{X}_1-\mathrm{X}_2)\widehat{\mathrm{Y}}}{p}\Big)\Big)\Big)\\ &= \mathbf{E}\Big(\Big|\mathbf{E}\Big(e\Big(\frac{\mathrm{X}\mathrm{X}_1\widehat{\mathrm{Y}}}{p}\Big)\Big)\Big|^2\Big) = \mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}_1\widehat{\mathrm{Y}}))|^2),\end{aligned}$$

leading to the conclusion since X and X_1 are identically distributed. □

Proof of Proposition 5.1. We define $\alpha \geqslant 1$ by $\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) = \alpha^{-1}$. By Lemma 6.5, we have then

$$\mathbf{E}(\varrho_{\mathrm{Y}}(\mathrm{XY})) = \frac{\varrho_{\mathrm{Y}}(0)}{\alpha}.$$

If the conditions (6.2) are not valid, then by construction this implies that the bound

$$\mathbf{E}(|\varphi_{\mathrm{X}}(\mathrm{X}\widehat{\mathrm{Y}})|^2) = \alpha^{-1} \leqslant 4(\varrho_{\mathrm{X}}(0) + \varrho_{\mathrm{Y}}(0))$$

holds. On the other hand, if these conditions are satisfied, then we can apply Proposition 6.1 to deduce the existence of $A \subset \mathbf{F}_p^\times$ with

$$\max(A+A, A\cdot A) \ll \alpha^d |A|$$

and

$$\frac{1}{\alpha^d \varrho_Y(0)} \ll |A| \ll \frac{\alpha}{\varrho_Y(0)},$$

where d and the implied constants are absolute (and explicit).

Let $\eta > 0$. We distinguish two further cases:

(1) If $|A| \leqslant p^{1-\eta}$, then denoting by $\delta > 0$ the exponent in Theorem 2.3 for $\gamma = \eta$, we have $\alpha^d \gg |A|^\delta$. It follows that $\alpha^d \gg \alpha^{-d\delta}\varrho_Y(0)^{-\delta}$, and hence

$$\mathbf{E}(|\varphi_X(X\widehat{Y})|^2) = \alpha^{-1} \ll \varrho_Y(0)^{\delta/(d+d\delta)}.$$

(2) If $|A| > p^{1-\eta}$, then

$$\mathbf{E}(|\varphi_X(X\widehat{Y})|^2) = \alpha^{-1} \ll \frac{1}{|A|\varrho_Y(0)} \ll \frac{p^{-1+\eta}}{\varrho_Y(0)}.$$

All three of the bounds thus obtained imply that the estimate (5.1) holds (with $\beta = \min(1, \delta/(d+d\delta))$), concluding the proof. □

We now come to the proof of Proposition 5.4. Only in the last step will the specific properties of the distribution of S be important.

Proof of Proposition 5.4. We recall the definition

$$X_k = \sum_{i=1}^{k} (S_{2i-1} - S_{2k}), \qquad k \geqslant 1,$$

of the random walk and the formula $\varphi_{X_k} = |\varphi_S|^{2k}$.

We observe first that for any integer $k \geqslant 1$ and $\nu > 0$, provided the condition $4k\nu \leqslant \theta$ is satisfied, the estimate

$$\varrho_{X_{2k}}(0) = \frac{M_{X_k}}{p} = \frac{1}{p}\sum_{a\in\mathbf{F}_p} |\varphi_S(a)|^{4k} \geqslant |\Lambda_\nu| p^{-1-\theta} \tag{6.8}$$

holds by (3.2) and the definition of Λ_ν.

We now claim that if p is large enough, depending only on θ, then we can find some integer $k \geqslant 1$ and $\nu < \frac{1}{2}\theta$, independent of p, such that $4k\nu \leqslant \theta$ and

$$p^{-\theta} \leqslant \frac{|\Lambda_\nu|}{M_{X_k}}, \tag{6.9}$$

which, together with (6.8) and the formula $\varrho_{X_{2k}}(0) = M_{X_k}/p$, ensures that (5.2) holds for these choices of k and ν.

To prove the claim, we first note that there is a general upper bound

$$M_{X_k} \leqslant |\Lambda_{1/k}| + p\cdot(p^{-4k})^k = |\Lambda_{1/k}| + p^{-3} \leqslant |\Lambda_{1/k}|(1+p^{-3}),$$

valid for any integer $k \geqslant 1$. Now, given $k \geqslant 1$, we denote $k_+ = \lceil \frac{\theta}{k^2} \rceil$. If the inequality $\mathrm{M}_{\mathrm{X}_k} > p^\theta |\Lambda_{1/k_+}|$ holds, then it follows that

$$|\Lambda_{1/k_+}| \leqslant |\Lambda_{1/k}| p^{-\theta}(1+p^{-3}).$$

Iterating this observation m times, starting from $k=4$, we see that *either* we find $k \geqslant 1$ such that (6.9) holds for $\nu = 1/k_+$, or we have

$$|\Lambda_{1/k}| \leqslant p^{1-m\theta}(1+p^{-3})^m$$

for $m \geqslant 1$ and some k depending on m. But for suitable m, we obtain $|\Lambda_{1/k}| < 1$, which is a contradiction since $0 \in \Lambda_\nu$ for all ν.

Our next goal is the inequality

$$\mathbf{E}(|\varphi_{\mathrm{X}_k}(a\mathrm{X}_k)|^2) \geqslant \varphi_{\mathrm{S}}(a)^{4k} \tag{6.10}$$

for all $k \geqslant 1$ and $a \in \mathbf{F}_p$, and this will depend on the specific choice of the random walk. Indeed, we first have

$$\mathbf{E}(|\varphi_{\mathrm{X}_k}(a\mathrm{X}_k)|^2) = \mathbf{E}(\varphi_{\mathrm{X}_k}(a\mathrm{X}_{2k})) = \mathbf{E}(|\varphi_{\mathrm{S}}(a\mathrm{X}_{2k})|^{2k}) \geqslant \mathbf{E}(\varphi_{\mathrm{S}}(a\mathrm{X}_{2k}))^{2k},$$

by Jensen's inequality. However, by a discrete Fubini, we have

$$\mathbf{E}(\varphi_{\mathrm{S}}(a\mathrm{X}_{2k})) = \mathbf{E}(|\varphi_{\mathrm{X}_k}(a\mathrm{S})|^2)$$

and $\mathbf{E}(|\varphi_{\mathrm{X}_k}(a\mathrm{S})|^2) = \varphi_{\mathrm{X}_k}(a)^2$ since $\varphi_{\mathrm{X}_k}(a\mathrm{S}) = \varphi_{\mathrm{X}_k}(a)$ (the crucial fact from Lemma 5.3), which gives (6.10).

We can then finally deduce (5.3). From (6.9) and the condition $4k\nu \leqslant \theta$, we deduce the lower bound

$$\mathbf{P}(\widehat{\mathrm{X}}_{2k} \in \Lambda_\nu) \geqslant p^{-\theta} \frac{|\Lambda_\nu|}{\mathrm{M}_{\mathrm{X}_k}} \geqslant p^{-2\theta},$$

and then from (6.10), we get

$$\mathbf{E}(|\varphi_{\mathrm{X}_k}(\mathrm{X}_k\widehat{\mathrm{X}}_{2k})|^2) \geqslant \mathbf{E}(\varphi_{\mathrm{X}_k}(\widehat{\mathrm{X}}_{2k})^{4k}) \geqslant p^{-4k^2\nu}\mathbf{P}(\widehat{\mathrm{X}}_{2k} \in \Lambda_\nu) \geqslant p^{-4k^2\nu-2\theta} \geqslant p^{-10\theta}.$$

□

7. Remarks

We conclude with a few brief remarks.

(1) One interpretation of Theorem 1.1 is that it is one more avatar of the fact that the additive and multiplicative structures of a finite field (or of the integers) are fairly "independent": it concerns the *additive* Fourier transform of a *multiplicative* subgroup. In this sense, it is of a flavor comparable with the sum-product theorem.

One may however then wonder about exchanging the role of addition and multiplication. And whereas the sum-product theorem is fully symmetric, the "dual" of Theorem 1.1 would become the problem of estimating sums of *multiplicative* (Dirichlet) characters modulo p over *very short* intervals in $\mathbf{F}_p$ – a problem which is intimately related with the Generalized Riemann Hypothesis and properties of Dirichlet L-functions. (We see short intervals as analogues of small multiplicative subgroups in view of their additive properties, which

makes them behave quite similarly to non-existent small additive subgroups; this is reasonable especially because Theorem 1.1 does extend to geometric progressions in addition to multiplicative subgroups.)

Could the proof of Theorem 1.1 give insight about such character sums? This doesn't seem to be likely, because there is no analogue of Lemma 5.3 (e.g., the existence of *one* large character sum for a non-trivial character does not, a priori, lead to the existence of any other). Ultimately, this reflects the fact that addition and multiplication *are not* symmetric in the definition of a field: multiplication is distributive with respect to addition, and not the opposite, so that multiplication by non-zero elements give automorphisms of the *additive* group of a field, leading to symmetry properties of the *additive* Fourier transform of multiplicative subgroups.

(2) One can also ask if there are echoes in this proof of more classical ideas in the study of exponential sums (such as those of Weyl, van der Corput and Vinogradov, see e.g. [7, Ch. 8]).

We see at least two clear links of this type:

- The use of $|\varphi_S|^2$ and higher powers is very much in the spirit of "creating new points of summation" or Weyl differencing.
- The link in Lemma 6.5, based on harmonic analysis, between averages of the Fourier transform and averages of the "density" ϱ_Y is an example of reduction of averages of exponential sums to point counting.

One related remark is that if we consider, instead of the crucial expression $\mathbf{E}(|\varphi_X(X\widehat{Y})|^2)$ in Proposition 5.1, the simpler $\mathbf{E}(|\varphi_X(\widehat{Y})|^2)$, then we get (up to normalization) simply the fourth moment of $\varphi_X(a)$, instead of a kind of average "twisted" fourth moment.

(3) Another parallel is with the work of Bourgain and Gamburd [5] on expansion properties of Cayley graphs of $\mathbf{SL}_2(\mathbf{F}_p)$, which is almost contemporary with Theorem 1.1. For instance, the crucial "L^2-flattening lemma" of Bourgain and Gamburd [5, Prop. 2] can be interpreted as a quantitative statement of decay of $\mathbf{P}(Y = 0)$ for a stepping Y of certain random variables X on $\mathbf{SL}_2(\mathbf{F}_p)$. Lemma 5.3 also has a similar flavor to the use of the "pseudo-randomness" of $\mathbf{SL}_2(\mathbf{F}_p)$ (i.e., the absence of non-trivial irreducible representations of small dimension) in [5, Prop. 1].

References

[1] J. Bourgain, N.H. Katz and T. Tao: *A sum-product estimate in finite fields, and applications*, GAFA 14 (2004), 27–57.

[2] J. Bourgain: *Exponential sums, equidistribution and pseudo-randomness*, talk at I.A.S, December 3, 2008; https://www.youtube.com/watch?v=s1EhZQ5kSNw.

[3] J. Bourgain: *Mordell's exponential sums estimate revisited*, Journal A.M.S. 18 (2005), 477–499.

[4] J. Bourgain, A.A. Glibichuk and S. 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] J. Bourgain and A. Gamburd: *Uniform expansion bounds for Cayley graphs of* $\mathbf{SL}_2(\mathbf{F}_p)$, Ann. of Math. 167 (2008), 625–642.

[6] E. Breuillard: *A brief introduction to approximate groups*, in "Thin groups and super-strong approximation", edited by E. Breuillard and H. Oh, MSRI Publications Vol. 61, Cambridge Univ. Press, 2014.

[7] H. Iwaniec and E. Kowalski: *Analytic Number Theory*, Colloquium Publ. 53, A.M.S, 2004.

[8] E. Kowalski: *Introduction to additive combinatorics*, ETH lecture notes (2023); https://www.math.ethz.ch/~kowalski/additive-combinatorics.pdf
[9] P. Kurlberg: *Bounds on exponential sums over small multiplicative subgroups*, in "Additive combinatorics", CRM Proc. Lecture Notes, 43, A.M.S, 2007.
[10] T. Schoen: *New bounds in Balog–Szemerédi–Gowers*, Combinatorica 35 (2015), 695–701.
[11] I. Shkredov: *Some remarks on the asymmetric sum-product phenomenon*, Moscow J. Comb. Number Th. 8 (2019), 15–41.
[12] I. Shparlinski: *Estimates for Gauss sums*, Mat. Zametki 50 (1991), 122–130.

(E. Kowalski) D-MATH, ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland
Email address: kowalski@math.ethz.ch