Forum of Mathematics, Sigma (2013), Vol. 1, e2, 40 pages
doi:10.1017/fms.2013.2

MIXING FOR PROGRESSIONS IN NONABELIAN GROUPS

TERENCE TAO
UCLA Department of Mathematics, Los Angeles, CA 90095-1555, USA;
email: tao@math.ucla.edu

Received 11 December 2012; accepted 31 May 2013

Abstract

We study the mixing properties of progressions (x, xg, xg^2), (x, xg, xg^2, xg^3) of length three and four in a model class of finite nonabelian groups, namely the special linear groups $\mathrm{SL}_d(F)$ over a finite field F, with d bounded. For length three progressions (x, xg, xg^2), we establish a strong mixing property (with an error term that decays polynomially in the order $|F|$ of F), which among other things counts the number of such progressions in any given dense subset A of $\mathrm{SL}_d(F)$, answering a question of Gowers for this class of groups. For length four progressions (x, xg, xg^2, xg^3), we establish a partial result in the $d = 2$ case if the shift g is restricted to be diagonalizable over F, although in this case we do not recover polynomial bounds in the error term. Our methods include the use of the Cauchy–Schwarz inequality, the abelian Fourier transform, the Lang–Weil bound for the number of points in an algebraic variety over a finite field, some algebraic geometry, and (in the case of length four progressions) the multidimensional Szemerédi theorem.

2010 Mathematics Subject Classification: 11B30, 20D60

1. Introduction

Let $G = (G, \cdot)$ be a finite group, not necessarily abelian. Given a natural number $k \geq 1$ and k functions $f_0, \ldots, f_{k-1} : G \to \mathbf{C}$, we define the k-linear form

$$\Lambda_{k,G}(f_0, \ldots, f_{k-1}) := \mathbf{E}_{x,g \in G} \prod_{i=0}^{k-1} f_i(xg^{i-1}),$$

where $\mathbf{E}$ denotes the averaging notation

$$\mathbf{E}_E f := \mathbf{E}_{x\in E} f(x) := \frac{1}{|E|}\sum_{x\in E} f(x)$$

for nonempty finite sets E and complex-valued functions f on E, with $|E|$ denoting the cardinality of the set E. Thus, for instance, if A is a subset of G, with the associated indicator function $1_A: G \to \{0, 1\}$, $\Lambda_{k,G}(1_A, \ldots, 1_A)$ denotes the number of (possibly degenerate) length k geometric progressions $(x, xg, \ldots, xg^{k-1})$ in A, divided by $|G|^k$.

The form $\Lambda_{k,G}$ is easily computed for $k = 1, 2$:

$$\Lambda_{1,G}(f_0) = \mathbf{E}_G f_0$$
$$\Lambda_{2,G}(f_0, f_1) = (\mathbf{E}_G f_0)(\mathbf{E}_G f_1).$$

Now we turn to the $k = 3$ case. If f_0, f_1, f_2 are selected in a sufficiently 'random' fashion, then probabilistic heuristics suggest that one has

$$\Lambda_{3,G}(f_0, f_1, f_2) \approx (\mathbf{E}_G f_0)(\mathbf{E}_G f_1)(\mathbf{E}_G f_2), \tag{1.1}$$

and, more generally,

$$\Lambda_{k,G}(f_0, \ldots, f_{k-1}) \approx \prod_{i=0}^{k-1} \mathbf{E}_G f_i. \tag{1.2}$$

However, if G has a nontrivial low-dimensional unitary representation $\rho: G \to U_d(\mathbf{C})$ for some small d, then it becomes possible to violate the heuristic (1.1). Indeed, if one lets B be a small neighbourhood of the identity in $U_d(\mathbf{C})$, and sets B' to be the slightly larger neighbourhood

$$B' := B \cdot B^{-1} \cdot B := \{b_1 b_2^{-1} b_3 : b_1, b_2, b_3 \in B\},$$

with the associated preimages $A := \rho^{-1}(B)$, $A' := \rho^{-1}(B')$, then from the identity

$$\rho(xg^2) = \rho(xg)\rho(x)^{-1}\rho(xg)$$

we see that $xg^2 \in B'$ whenever $x, xg \in B$. In particular, we have

$$\Lambda_{3,G}(1_A, 1_A, 1_{A'}) = \Lambda_{2,G}(1_A, 1_A) = (\mathbf{E}_G 1_A)(\mathbf{E}_G 1_A),$$

which violates (1.1) if B (and hence B') is small enough; if the dimension d is small, this can be done with a relatively large value for the density $\mathbf{E}_G 1_A$. A similar argument demonstrates a deviation from (1.2) for any $k \geq 3$.

The deviation from (1.1) is most pronounced in the case when G is abelian (in which case all irreducible unitary representations of G are in fact one dimensional). In this case, we will switch to additive notation, and write the

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

group operation of G as $+$, so that

$$\Lambda_{3,G}(f_0,f_1,f_2) := \mathbf{E}_{x,g\in G} f_0(x) f_1(x+g) f_2(x+2g). \tag{1.3}$$

The analysis of this form usually begins by introducing the Fourier transform

$$\hat{f}(\xi) := \mathbf{E}_{x\in G} f(x) e(-\xi\cdot x)$$

for all ξ in the Pontryagin dual $\hat{G}$ of G, defined as the space of all homomorphisms $\xi: x \mapsto \xi\cdot x$ from G to the (additive) unit circle $\mathbf{R}/\mathbf{Z}$, where $e(x) := e^{2\pi i x}$; of course, $\hat{G}$ is encoding the irreducible one-dimensional unitary representations of G mentioned previously. Using the Fourier inversion formula

$$f(x) := \sum_{\xi\in\hat{G}} \hat{f}(\xi) e(\xi\cdot x),$$

one soon arrives at the useful identity

$$\Lambda_{3,G}(f_0,f_1,f_2) = \sum_{\xi\in\hat{G}} \hat{f}_0(\xi)\hat{f}_1(-2\xi)\hat{f}_2(\xi)$$

relating the magnitude of $\Lambda_{3,G}(f_0,f_1,f_2)$ with the size of the Fourier coefficients of f_0,f_1,f_2. Note that the heuristic (1.1) corresponds to the $\xi = 0$ term in this sum; the point is that the nonzero frequencies $\xi \neq 0$ can also give a significant contribution.

Using the above identity, one can eventually establish the Roth-type theorem

$$\Lambda_{3,G}(1_A, 1_A, 1_A) \geq c_3(\delta) \tag{1.4}$$

for any $0 < \delta \leq 1$, any finite abelian group G, and any subset $A \subset G$ with $|A| \geq \delta|G|$, where $c_3(\delta) > 0$ depends only on δ; see, for example, [29, Theorem 10.9]. In a similar vein, we have the deep theorem of Szemerédi [26], which implies (strictly speaking, the original theorem of Szemerédi only treats the case when G is a cyclic group, but subsequent proofs of Szemerédi's theorem (such as the hypergraph-based proofs in [11, 22, 23, 27]) allow for one to handle arbitrary abelian groups G) the more general lower bound,

$$\Lambda_{k,G}(1_A, \ldots, 1_A) \geq c_k(\delta), \tag{1.5}$$

for all $k \geq 1$ and $0 < \delta \leq 1$, any finite abelian group G, and any $A \subset G$ with $|A| \geq \delta|G|$, where $c_k(\delta) > 0$ depends only on k and δ.

REMARK 1.1. More explicit bounds for $c_3(\delta)$ are known. For general abelian groups G, an argument of Bourgain [5] gives $c_3(\delta) \geq c\delta^{C/\delta^2}$ for some absolute constants $c, C > 0$; see, for example, [29, Theorem 10.30]. In the case when G is a cyclic group, the strongest bound to date is due to Sanders [24],

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

who (in our notation) established that $c_3(\delta) \geq c\delta^{C\log^4(1/\delta)/\delta}$; on the other hand, in this case one also has the upper bound $c_3(\delta) \leq C\delta^{c\log(1/\delta)}$ due to Behrend [3]. When G is a vector space over a fixed finite field F of odd order (such as F_3), the best bound is due to Bateman and Katz [2], who established $c_3(\delta) \geq \exp(-C\delta^{c-1})$ for some constants $C, c > 0$ depending only on F. For $k > 3$ and for cyclic groups, the explicit bounds known are weaker: for $k = 4$, the results in [13] give $c_4(\delta) \geq c\exp(-C\delta^{-C\log(1/\delta)})$, while, for higher k, the results in [10] give $c_k(\delta) \geq c_k \exp(\exp(-C_k\delta^{-C_k}))$ for some constants $c_k, C_k > 0$ depending on k; in the other direction, a modification of the Behrend construction [21] gives $c_k(\delta) \leq C_k\delta^{c_k\log^{c_k}(1/\delta)}$. For general groups, explicit lower bounds on $c_k(\delta)$ are known thanks to the recent quantitative work on the density Hales–Jewett theorem [19] or the hypergraph removal lemma [12, 22, 23, 27], but the bounds are rather poor.

Now we turn to the case when G is not necessarily abelian, and in particular in the *quasirandom* case in which G has no low-dimensional representations. More precisely, following Gowers [12], call a finite group *GD-quasirandom* if the only irreducible unitary representations $\rho: G \to U_d(\mathbf{C})$ have dimension d greater than or equal to D. A model example of quasirandom groups is provided by the special linear groups over a finite field, as in the following proposition.

PROPOSITION 1.2 (Quasirandomness of special linear group). *Let $d \geq 2$ be an integer, and let F be a finite field. Then the group $\mathrm{SL}_d(F)$ of $d \times d$ matrices with coefficients in F of determinant one is $c_d|F|^{d-1}$-quasirandom, for some $c_d > 0$ depending only on d.*

Proof. This follows from the results in [16]. The case when $d = 2$ and $|F|$ has prime order is classical, dating back to the work of Frobenius. Similar results hold for other finite (almost) simple groups of Lie type and bounded rank; see [16]. □

When D is large, one expects better mixing properties in the forms $\Lambda_{k,G}$. To illustrate this, we introduce the variant expression

$$\Lambda^*_{k,G}(f_0, \ldots, f_{k-1}) := \mathbf{E}_{g\in G}\left|\mathbf{E}_{x\in G}\prod_{i=0}^{k-1} f_i(xg^{i-1}) - \prod_{i=0}^{k-1}\mathbf{E}_G f_i\right|,$$

which controls the number of length k progressions for a *single* (generic) shift g, as opposed to the average number over all such g. This expression vanishes for $k = 1$, but can be nontrivial for $k > 1$. From the triangle inequality, we have

$$\left|\Lambda_{k,G}(f_0, \ldots, f_{k-1}) - \prod_{i=0}^{k-1}\mathbf{E}_G f_i\right| \leq \Lambda^*_{k,G}(f_0, \ldots, f_{k-1}), \tag{1.6}$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

and so the heuristic (1.2) holds whenever $\Lambda^*_{k,G}(f_0,\dots,f_{k-1})$ is small. However, when one has a low-dimensional representation $\rho: G \to U_d(\mathbf{C})$, it is possible for $\Lambda^*_{k,G}(f_0,\dots,f_{k-1})$ to be large even when (1.2) holds. Consider for instance the $k=2$ case, in which (1.2) holds exactly. If we let B be a small neighbourhood of the identity in $U_d(\mathbf{C})$ with preimage $A := \rho^{-1}(B)$ as before, and set $A' := \rho^{-1}(B^{-1}\cdot B)$, we see that $1_A(x)1_A(xg)$ vanishes whenever $g \notin A'$, and thus

$$\Lambda^*_{2,G}(1_A, 1_A) = \mathbf{E}_{g\in G}|\mathbf{E}_{x\in G}1_A(x)1_A(xg) - (\mathbf{E}_G 1_A)^2|$$

can be lower bounded by $(\mathbf{E}_G 1_A)^2(1 - \mathbf{E}_G 1_{A'})$, which can be somewhat large if B is chosen small enough, and d is small.

As observed first by Gowers [12], though, $\Lambda^*_{2,G}$ becomes much smaller in the quasirandom case. This is elegantly captured by the inequality

$$\|f_1 * f_2\|_{L^2(G)} \le D^{-1/2}|G|\|f_1\|_{L^2(G)}\|f_2\|_{L^2(G)} \tag{1.7}$$

of Babai *et al*. [1], for any D-quasirandom group G and any functions $f_1, f_2: G \to \mathbf{C}$ with at least one of f_1, f_2 having mean zero, where

$$\|f\|_{L^2(G)} := (\mathbf{E}_{x\in G}|f(x)|^2)^{1/2}$$

and $*$ denotes the discrete (ordinarily, one would normalize this convolution by $1/|G|$ for compatibility with the averaging in the $L^2(G)$ norm, but it will be convenient to use the discrete normalization because we will be passing from a group G to various subgroups of G in subsequent arguments) convolution

$$f_1 * f_2(x) := \sum_{y\in G} f_1(y)f_2(y^{-1}x) = \sum_{y\in G} f_1(xy^{-1})f_2(y);$$

see [1] or [4, Proposition 3]. Note that (1.7) is an improvement by a factor of $D^{-1/2}$ over the trivial bound of $|G|\|f_1\|_{L^2(G)}\|f_2\|_{L^2(G)}$ arising from the Young and Cauchy–Schwarz inequalities.

The estimate (1.7) has the following useful corollary.

LEMMA 1.3 ($k=2$ mixing for quasirandom groups). *If G is a D-quasirandom group, then*

$$\Lambda^*_{2,G}(f_1, f_2) \le D^{-1/2}\|f_1\|_{L^2(G)}\|f_2\|_{L^2(G)}.$$

Proof. Observe that the expression $\Lambda^*_{2,G}(f_1, f_2)$ does not change if f_1 or f_2 is modified by an additive constant. Thus we may normalize f_1 and f_2 to both have mean zero. We can then write

$$\Lambda^*_{2,G}(f_1, f_2) = \mathbf{E}_{g\in G} f_0(g)\mathbf{E}_{x\in G} f_1(x)f_2(xg)$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

for some function $f_0: G \to \mathbf{C}$ of magnitude 1. The right-hand side can be rewritten, after a change of variables, as

$$\frac{1}{|G|}\mathbf{E}_{y\in G}(f_0 * f_1)(y)f_2(y).$$

The claim then follows from (1.7) and the Cauchy–Schwarz inequality. □

In [12], Gowers posed the question of whether results such as Lemma 1.3 could be extended to higher values of k, so that the heuristic (1.1) or (1.2) could hold for sufficiently quasirandom groups. We were not able to settle this question in general, but in the $k = 3$ case we can affirmatively answer the question for a model class of quasirandom groups, namely the special linear groups $\mathrm{SL}_d(F)$ over a finite field F.

THEOREM 1.4. *Let F be a finite field, and set $G := \mathrm{SL}_d(F)$ for some $d \geq 2$. Then we have*

$$|\Lambda^*_{3,G}(f_0, f_1, f_2)| \ll_d |F|^{-\min(d-1,2)/8} \prod_{i=0}^{2} \|f_i\|_{L^\infty(G)}$$

for all functions $f_0, f_1, f_2: G \to \mathbf{C}$, where $\|f\|_{L^\infty(G)} := \sup_{x\in G} |f(x)|$. Here and in what follows we use $Y \ll_d X$, $X \gg_d Y$, or $Y = O_d(X)$ to denote the estimate $|Y| \leq C_d X$ for some C_d depending only on d, and similarly with d replaced by other sets of parameters. In particular, from (1.6), one has

$$\Lambda_{3,G}(f_0, f_1, f_2) = (\mathbf{E}_G f_0)(\mathbf{E}_G f_1)(\mathbf{E}_G f_2) + O_d\left(|F|^{-\min(d-1,2)/8} \prod_{i=0}^{2} \|f_i\|_{L^\infty(G)}\right).$$

Theorem 1.4 is proven primarily through application of the Cauchy–Schwarz inequality and Lemma 1.3; we give this proof in Sections 2–4. The key point is that the nonabelian nature of G means that the application of the Cauchy–Schwarz inequality creates more averaging than is seen in the abelian case. The exponent $\min(d-1, 2)/8$ is unlikely to be optimal. By taking f_0, f_1, f_2 to be constant on left cosets gH of a proper subgroup H and of mean zero, we see that one cannot replace the quantity $|F|^{-\min(d-1,2)/8}$ by anything much smaller than $|H|/|G|$; in particular, if we take H to be the Borel subgroup of upper-triangular matrices in G, we see that one cannot replace $\min(d-1, 2)/8$ by any exponent greater than $(d(d-1))/2$. It is likely that one can extend Theorem 1.4 to other finite simple groups (to be pedantic, $\mathrm{SL}_d(F)$ is usually not a simple group, due to its nontrivial centre; but it is a bounded cover of a finite simple group, namely $\mathrm{PSL}_d(F)$. Note that the results for $\mathrm{SL}_d(F)$ in this paper automatically descend to the quotient group $\mathrm{PSL}_d(F)$ without difficulty) of Lie type with bounded rank, but we will not do so here.

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

Applying Theorem 1.4 to indicator functions $f_0 = 1_A, f_1 = 1_B, f_2 = 1_C$, and using Markov's inequality, we obtain in particular the 'weak mixing' bound

$$\mu(A \cap Bg \cap Cg^2) = \mu(A)\mu(B)\mu(C) + O_d(|F|^{-\min(d-1,2)/16})$$

for a proportion $1 - O_d(|F|^{-\min(d-1,2)/16})$ of $g \in G$, where $\mu(A) := \mathbf{E}_G 1_A = |A|/|G|$ denotes the density of A in G.

We conjecture that Theorem 1.4 can be extended to higher values of k than $k = 3$ (possibly with a smaller exponent than $\min(d-1, 2)/8$). Unfortunately, the Cauchy–Schwarz argument does not seem to extend beyond $k = 3$; in contrast to the abelian case, in the nonabelian setting it appears that when $k > 3$, each application of the Cauchy–Schwarz inequality *increases* the complexity of the resulting form, rather than decreasing it as in the abelian case. However, we are able to establish the following weak partial result in the $k = 4$, $d = 2$ case, in which the shift g is restricted to be diagonalizable.

THEOREM 1.5. *Let F be a finite field, and set $G := \mathrm{SL}_2(F)$. Let S denote all the elements of G which are diagonalizable over F. Then, for all functions $f_0, f_1, f_2, f_3 : G \to \mathbf{C}$, one has*

$$\mathbf{E}_{g \in S} \left| \mathbf{E}_{x \in G} \prod_{i=0}^{3} f_i(xg^{i-1}) - \prod_{i=0}^{k-1} \mathbf{E}_G f_i \right| = o_{|F| \to \infty}\left(\prod_{i=0}^{3} \|f_i\|_{L^\infty(G)} \right),$$

where $o_{|F| \to \infty}(X)$ denotes a quantity bounded by $c(|F|)X$ for some quantity $c(|F|)$ that goes to zero as $|F|$ goes to infinity.

It is easy to show that, for large $|F|$, S has density about $1/2$ in G; see Section 6. The main reason why the shift g is restricted to S in our arguments is in order to ensure that g is contained in a nontrivial metabelian subgroup of G; for instance, if g is a diagonal matrix with entries in F, then it is contained in the Borel subgroup B of upper-triangular matrices in G. The argument is rather *ad hoc* in nature, combining the Cauchy–Schwarz inequality and the abelian Fourier transform with some explicit nonabelian effects coming from the algebraic structure of progressions in the Borel group. It also relies on (a quantitative version of) the multidimensional Szemerédi theorem of Furstenberg and Katznelson [8], which is the reason for the poor decay in $|F|$. Finally, to pass from the Borel subgroup back to the full group, an expansion result in $\mathrm{SL}_2(F)$, related to the Bourgain–Gamburd expansion theory in this group, is also required.

REMARK 1.6. The results in this paper concern the mixing properties of the patterns (x, xg, xg^2) and (x, xg, xg^2, xg^3) for an explicit class of quasirandom groups, namely the special linear groups. In a recent paper with Bergelson [4],

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

we also establish some mixing properties for the patterns (x, xg, gx) and (g, x, xg, gx) in arbitrary quasirandom groups. While the end results of both papers are superficially similar in nature, the proof techniques turn out to be completely different, with the results in [4] relying on nonstandard analysis, the triangle removal lemma from graph theory, and ergodic theorems involving idempotent ultrafilters. In both cases, the methods are tailored to the specific patterns being counted, and it appears we are still quite far from a general theory that can cover all nonabelian patterns involving two or more variables such as x, g.

We also remark that, in [28], some mixing properties of patterns of the form $(x, y, P(x, y))$ were established when $P: G \times G \to G$ was a definable function over a finite field of large characteristic. However, the arguments in that paper (which also involve the Cauchy–Schwarz inequality, but applied in a slightly different fashion) required $\{(P(x, y), P(x, y'), P(x', y), P(x', y')) : x, y \in G\}$ to be sufficiently Zariski dense in G^4. This is not the case for the pattern (x, xg, xg^2) (in which $P(x, y) := yx^{-1}y$), since $P(x, y)$ and $P(x, y')$ are necessarily conjugate to each other.

2. A general bound for $\Lambda_{3,G}$

Let us define the *reduced spectral norm* $\|\mu\|_{S(G)}$ of a function $\mu: G \to \mathbf{C}$ to be the best constant such that

$$\|f * \mu\|_{L^2(G)} \leq \|\mu\|_{S(G)} \|f\|_{L^2(G)} \tag{2.1}$$

whenever $f: G \to \mathbf{C}$ has mean zero; thus

$$|\mathbf{E}_{z \in G} f_1(z)(f_2 * \mu)(z)| \leq \|\mu\|_{S(G)} \|f_1\|_{L^2(G)} \|f_2\|_{L^2(G)} \tag{2.2}$$

for all $f_1, f_2: G \to \mathbf{C}$, as can be seen by splitting f_1, f_2 into constant and mean zero components, and noting that all cross terms vanish.

REMARK 2.1. From the Peter–Weyl theorem, one can also write $\|\mu\|_{S(G)}$ as

$$\|\mu\|_{S(G)} = \sup_{\rho} \left\| \sum_{g \in G} \mu(g) \rho(g) \right\|_{\mathrm{op}},$$

where $\rho: G \to U(V)$ ranges over all nontrivial irreducible finite-dimensional unitary representations of G. We will not make much use of this representation-theoretic interpretation of the reduced spectral norm here, although we remark that this interpretation can be used to derive the basic quasirandomness inequality (1.7) (or (2.4) below).

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

The reduced spectral norm $\|\mu\|_{S(G)}$ is clearly a seminorm, and in particular it obeys the triangle inequality. From Minkowski's inequality, we have the crude bound

$$\|\mu\|_{S(G)} \leq \|\mu\|_{\ell^1(G)}. \tag{2.3}$$

From (1.7), we also have the more refined estimate

$$\|\mu\|_{S(G)} \leq D^{-1/2}|G|^{1/2}\|\mu\|_{\ell^2(G)} \tag{2.4}$$

when G is D-quasirandom. If we split μ into the region where $\mu(x) > C_0/|G|$, and the region where $\mu(x) \leq C_0/|G|$, for some threshold $C_0 > 0$, and apply (2.3) to the latter and (2.4) to the former, we conclude that

$$\|\mu\|_{S(G)} \leq C_0 D^{-1/2} + \sum_{x\in G:\mu(x)>C_0/|G|} \mu(x). \tag{2.5}$$

By combining these estimates with the Cauchy–Schwarz inequality, we can obtain the following general bound on the quantity $\Lambda_3(f_0, f_1, f_2)$.

PROPOSITION 2.2. *Let $G = (G, \cdot)$ be a D-quasirandom group for some $D \geq 1$. Let $C_0 \geq 1$ be a parameter. Then we have*

$$\Lambda^*_{3,G}(f_0, f_1, f_2) \ll \left(C_0 D^{-1/2} + \mathbf{E}_{b,h\in G} \sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|} \mu_{b,h}(y)\right)^{1/4} \prod_{i=0}^{2} \|f_i\|_{L^\infty(G)} \tag{2.6}$$

for all functions $f_0, f_1, f_2 \colon G \to \mathbf{C}$, where, for each $b, h \in G$, $\mu_{b,h} \colon G \to \mathbf{C}$ is the function

$$\mu_{b,h} := \mathbf{E}_{g\in G}\mathbf{E}_{c\in Z(b)}\delta_{gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1}}, \tag{2.7}$$

where $Z(b) := \{c \in G : cb = bc\}$ is the centralizer of b.

One can view $\mu_{b,h}$ as a probability measure on G, describing the distribution of the random variable $gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1}$ when g is a randomly chosen element of G and c is a random element commuting with b. The estimate (2.6) becomes useful when $\mu_{b,h}$ is approximately uniformly distributed over G for typical b, h, so that $\sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|} \mu_{b,h}(y)$ is small.

Proof. When f_0 is equal to a constant c, we have

$$\Lambda^*_{3,G}(f_0, f_1, f_2) = |c|\Lambda^*_{2,G}(f_1, f_2),$$

and the claim then follows from Lemma 1.3. As $\Lambda^*_{3,G}$ is sublinear in each of the three arguments, we may thus assume that f_0 has mean zero. We then also assume

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

that f_0, f_1, f_2 are real valued, and normalize so that

$$\|f_0\|_{L^\infty(G)} = \|f_1\|_{L^\infty(G)} = \|f_2\|_{L^\infty(G)} = 1.$$

Our task is now to show that

$$|\Lambda^*_{3,G}(f_0, f_1, f_2)|^4 \ll C_0 D^{-1/2} + \mathbf{E}_{b,h\in G} \sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|} \mu_{b,h}(y).$$

Ever since the work of Gowers [9], it has been common to control expressions such as $\Lambda^*_{3,G}(f_0, f_1, f_2)$ via the Cauchy–Schwarz inequality. In the literature, this was mostly performed in the abelian case, but one can obtain a useful estimate via the Cauchy–Schwarz inequality in the nonabelian case too. First, we shift x by g^{-1} to obtain

$$\Lambda^*_{3,G}(f_0, f_1, f_2) = \mathbf{E}_{g\in G}|\mathbf{E}_{x\in G} f_0(xg^{-1}) f_1(x) f_2(xg)|$$

which we expand as

$$\Lambda^*_{3,G}(f_0, f_1, f_2) = \mathbf{E}_{x\in G} f_1(x)(\mathbf{E}_{g\in G} f_0(xg^{-1}) f_2(xg) f_3(g))$$

for some (if one is only interested in bounding $\Lambda_{3,G}(f_0, f_1, f_2)$ rather than $\Lambda^*_{3,G}(f_0, f_1, f_2)$, one can take $f_3 \equiv 1$, and the reader may wish to do so initially in the argument that follows in order to simplify the exposition) function $f_3 : G \to \mathbf{C}$ bounded in magnitude by 1. Applying the Cauchy–Schwarz inequality in x to eliminate f_1, we obtain

$$\Lambda^*_{3,G}(f_0, f_1, f_2) \leq (\mathbf{E}_{x\in G}|\mathbf{E}_{g\in G} f_0(xg^{-1}) f_2(xg) f_3(g)|^2)^{1/2}.$$

We can expand the right-hand side as

$$(\mathbf{E}_{x,g,g'\in G} f_0(xg^{-1}) f_0(x(g')^{-1}) f_2(xg) f_2(xg') f_3(g) f_3(g'))^{1/2}.$$

Making the change of variables $(y, g, a) := (xg, g, g^{-1}g')$, this becomes

$$(\mathbf{E}_{y,g,a\in G} f_0(yg^{-2}) f_0(yg^{-1}a^{-1}g^{-1}) f_2(y) f_2(ya) f_3(g) f_3(ga))^{1/2}.$$

If we define $\Delta_a f(y) := f(y)f(ya)$, this becomes

$$(\mathbf{E}_{y,a\in G} \Delta_a f_2(y)(\mathbf{E}_{g\in G} \Delta_{ga^{-1}g^{-1}} f_0(yg^{-2}) \Delta_a f_3(g)))^{1/2}.$$

Applying the Cauchy–Schwarz inequality in y, a to eliminate $\Delta_a f_2$, we thus have

$$\Lambda^*_{3,G}(f_0, f_1, f_2) \leq (\mathbf{E}_{y,a\in G}|\mathbf{E}_{g\in G} \Delta_{ga^{-1}g^{-1}} f_0(yg^{-2}) \Delta_a f_3(g)|^2)^{1/4}.$$

The right-hand side can be expanded as

$$(\mathbf{E}_{y,a,g,g'\in G} \Delta_{ga^{-1}g^{-1}} f_0(yg^{-2}) \Delta_{g'a^{-1}(g')^{-1}} f_0(y(g')^{-2}) \Delta_a f_3(g) \Delta_a f_3(g'))^{1/4}.$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

Making the change of variables $(z, b, g, h) := (yg^{-2}, ga^{-1}g^{-1}, g, g'g^{-1})$, we conclude the inequality

$$\begin{aligned}|\Lambda_{3,G}(f_0,f_1,f_2)| \leq (&\mathbf{E}_{z,b,g,h\in G}\Delta_b f_0(z)\Delta_{hbh^{-1}}f_0(zgh^{-1}g^{-1}h^{-1})\\ &\times \Delta_{g^{-1}b^{-1}g}f_3(g)\Delta_{g^{-1}b^{-1}g}f_3(hg))^{1/4}.\end{aligned} \tag{2.8}$$

The right-hand side of (2.8) can be viewed as a twisted weighted variant (indeed, in the model case when $f_3 \equiv 1$ and G is abelian, the right-hand side simplifies to $(\mathbf{E}_{z,b,h\in G}\Delta_b f_0(z)\Delta_b f_0(zh^{-2}))^{1/4}$, which (in the case that G has odd order) is precisely the Gowers norm $\|f_0\|_{U^2(G)}$) of the Gowers U^2 norm [9]. To control it, we begin by observing the self-averaging identity

$$\mathbf{E}_{h\in G}F(h) = \mathbf{E}_{h\in G}\mathbf{E}_{c\in C}F(hc)$$

for any nonempty set C and any function $F: G \to \mathbf{C}$. We apply this identity with C equal to the centralizer $Z(b) := \{c \in G : cb = bc\}$ of b and F equal to the expression being averaged on the right-hand side of (2.8); the point of this averaging is to exploit the trivial observation that the function $\Delta_{hbh^{-1}}f_0$ does not change if one replaces h by hc for an arbitrary $c \in Z(b)$. We conclude that

$$\begin{aligned}|\Lambda_{3,G}(f_0,f_1,f_2)| \leq (&\mathbf{E}_{z,b,g,h\in G}\mathbf{E}_{c\in Z(b)}\Delta_b f_0(z)\Delta_{hbh^{-1}}f_0(zgc^{-1}h^{-1}g^{-1}c^{-1}h^{-1})\\ &\times \Delta_{g^{-1}b^{-1}g}f_3(g)\Delta_{g^{-1}b^{-1}g}f_3(hcg))^{1/4}.\end{aligned}$$

We can rewrite the right-hand side as

$$|\mathbf{E}_{b,h\in G}\mathbf{E}_{z\in G}\Delta_b f_0(z)(\Delta_{hbh^{-1}}f_0 * \tilde{\mu}_{b,h})(z)|^{1/4}, \tag{2.9}$$

where $\tilde{\mu}_{b,h}$ is a weighted version (Returning to the model case when $f_3 \equiv 1$ and G is an abelian group of odd order, we have in this case that $\tilde{\mu}_{b,h} \equiv 1/|G|$, and (2.9) is again just the Gowers norm $\|f_0\|_{U^2(G)}$. The point is that for certain nonabelian groups G, one can still obtain some sort of equidistribution control on $\tilde{\mu}_{b,h}$ that makes it behave roughly like the uniform distribution $1/|G|$.) of $\mu_{b,h}$:

$$\tilde{\mu}_{b,h} := \mathbf{E}_{g\in G}\mathbf{E}_{c\in Z(b)}\delta_{gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1}}\Delta_{g^{-1}b^{-1}g}f_3(g)\Delta_{g^{-1}b^{-1}g}f_3(hcg).$$

Our task is now to show that

$$\begin{aligned}&|\mathbf{E}_{b,h\in G}\mathbf{E}_{z\in G}\Delta_b f_0(z)(\Delta_{hbh^{-1}}f_0 * \tilde{\mu}_{b,h})(z)|\\ &\ll C_0D^{-1/2} + \mathbf{E}_{b,h\in G}\sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|}\mu_{b,h}(y).\end{aligned} \tag{2.10}$$

From (2.2), we see that

$$|\mathbf{E}_{z\in G}\Delta_b f_0(z)(\Delta_{hbh^{-1}}f_0 * \tilde{\mu}_{b,h})(z)| \leq \|\tilde{\mu}_{b,h}\|_{S(G)} + |\mathbf{E}_{z\in G}\Delta_b f_0(z)|$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

(by splitting $\Delta_h f_0$ into constant and mean zero components). We may thus upper bound the left-hand side of (2.10) by

$$\mathbf{E}_{b,h\in G}\|\tilde{\mu}_{b,h}\|_{S(G)} + \mathbf{E}_{b\in G}|\mathbf{E}_{z\in G}\Delta_h f_0(z)|.$$

The second term is equal to $\Lambda^*_{2,G}(f_0, f_0)$, which by Lemma 1.3 is bounded by $D^{-1/2}$. As for the first term, we see from (2.5) and the pointwise bound $|\tilde{\mu}_{b,h}(x)| \leq \mu_{b,h}(x)$ that

$$\|\tilde{\mu}_{b,h}\|_{S(G)} \leq C_0 D^{-1/2} + \mathbf{E}_{b,h\in G} \sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|} \mu_{b,h}(y)$$

for each b, h. The claim follows. □

3. The case of SL_2

We can now establish the $d = 2$ case of Theorem 1.4, which serves as a simplified model for the general d case. From Propositions 1.2 and 2.2, it will suffice to show that

$$\mathbf{E}_{b,h\in G} \sum_{y\in G:\mu_{b,h}(y)\geq C_0/|G|} \mu_{b,h}(y) \ll |F|^{-1} \tag{3.1}$$

for some absolute constant $C_0 \geq 1$, where $\mu_{b,h}$ was defined in (2.7).

We now need to understand the distribution of $\mu_{b,h}$. Call an element b of $\mathrm{SL}_2(F)$ *regular semisimple* if its two eigenvalues (in the algebraic closure $\overline{F}$) are distinct, or equivalently if trace $b \neq \pm 2$. It is easy to see that all but $O(|F|^2)$ elements of G are regular semisimple. Since G has cardinality comparable to $|F|^3$, and each of the $\mu_{b,h}$ is normalized in ℓ^1, we thus see that the contribution of the nonregular semisimple b to (3.1) is $O(|F|^{-1})$, which is acceptable. Thus we may restrict attention to the regular semisimple b.

Now we study the quantity $\mu_{b,h}(y)$. It is a classical fact that $|F| \ll |Z(b)| \ll |F|$ (this also follows from the Lang–Weil bound, Proposition A.3). As such, we have

$$\mu_{b,h}(y) \ll |F|^{-4}|\{(g, c) \in G \times Z(b) : gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1} = y\}|,$$

which we rewrite as

$$\mu_{b,h}(y) \ll |F|^{-4}|\{(g, c) \in G \times Z(b) : gc^{-1}h^{-1}g^{-1} = yhc\}|.$$

If $c^{-1}h^{-1}$ is central (that is equal to ± 1), then $y = 1$, and the contribution to $\mu_{b,h}(1)$ of this case is $O(|F|^{-1})$. Now we consider the contribution of those c for which $c^{-1}h^{-1}$ is not central. Then the centralizer of $c^{-1}h^{-1}$ has cardinality $\gg |F|$,

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

and so every element k of $\mathrm{SL}_2(F)$ of the same trace as $c^{-1}h^{-1}$ has $O(|F|)$ representations of the form $gc^{-1}h^{-1}g^{-1}$. Of course, if k does not have the same trace as $c^{-1}h^{-1}$, it has no such representations. We conclude that

$$\mu_{b,h}(y) \ll |F|^{-1}\delta_{y=1} + |F|^{-3}|\{c \in Z(b) : \mathrm{trace}(yhc) = \mathrm{trace}(c^{-1}h^{-1})\}|.$$

For $a \in \mathrm{SL}_2(F)$, we see from direct computation (or the Cayley–Hamilton theorem) that $\mathrm{trace}(a^{-1}) = \mathrm{trace}(a)$. We thus have $\mu_{b,h}(y) \ll |F|^{-1}$ for $y = 1$, and for $y \neq 1$ we have

$$\mu_{b,h}(y) \ll |F|^{-3}|\{c \in Z(b) : \mathrm{trace}(yhc) = \mathrm{trace}(hc)\}|.$$

The centralizer $Z(b)$ is the set of F-points of the algebraic variety $\overline{Z(b)} := \{c \in \mathrm{SL}_2(\overline{F}) : cb = bc\}$, which is a curve of complexity (the complexity of an algebraic variety is defined in Definition A.1) $O(1)$. From Bezout's theorem, we conclude that the quantity $|\{c \in Z(b) : \mathrm{trace}(yhc) = \mathrm{trace}(hc)\}|$ is bounded by $O(1)$ unless the equation $\mathrm{trace}(yhc) = \mathrm{trace}(hc)$ holds for all $c \in \overline{Z(b)}$, in which case this quantity is bounded instead by $|F|$. For C_0 a sufficiently large absolute constant, we thus have

$$\sum_{y \in G : \mu_{b,h}(y) \geq C_0/|G|} \mu_{b,h}(y) \ll |F|^{-1} + |F|^{-2}|Y_{b,h}|,$$

where $Y_{b,h}$ is the set of all $y \in G$ such that $\mathrm{trace}(yhc) = \mathrm{trace}(hc)$ for all $c \in \overline{Z(b)}$. It will thus suffice to show that

$$|Y_{b,h}| \ll |F|$$

whenever b is regular semisimple.

Fix such a b. We may find a basis of $\overline{F}^2$ over $\overline{F}$ that makes b diagonal. As b is also regular semisimple, we conclude that

$$\overline{Z(b)} = \left\{ \begin{pmatrix} t & 0 \\ 0 & t^{-1} \end{pmatrix} : t \in \overline{F} \backslash 0 \right\}$$

in this basis, and so the constraint $\mathrm{trace}(yhc) = \mathrm{trace}(hc)$ for all $c \in \overline{Z(b)}$ is equivalent to the requirement that $yh - h$ vanishes on the diagonal. This constrains $Y_{b,h}$ to a two-dimensional subspace of the four-dimensional vector space $\mathrm{Mat}_{2\times 2}(\overline{F})$ of 2×2 matrices; as y also needs to have determinant 1, we conclude that $Y_{b,h}$ is constrained to a complexity $O(1)$ curve in this plane. By the Schwarz–Zippel lemma (see Proposition A.2), we conclude that $|Y_{b,h}| \ll |F|$, as required.

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

4. The case of SL_d

Now we turn to the general case of Theorem 1.4. This will basically be a reprise of the arguments in the preceding section, but with a heavier reliance on algebraic geometry in place of *ad hoc* computations.

We allow all implied constants to depend on d. As before, by Propositions 1.2 and 2.2, it suffices to establish the bound (3.1). We may assume that $|F|$ is sufficiently large depending on d, as the claim is trivial otherwise.

Again, call $b \in \mathrm{SL}_d(F)$ *regular semisimple* if it is diagonalizable in $\overline{F}$ with distinct eigenvalues. A well-known computation gives

$$|\mathrm{GL}_d(F)| = \prod_{i=0}^{d-1}(|F|^d - |F|^i) = (1 + O(|F|^{-1}))|F|^{d^2};$$

since $|G| = |\mathrm{GL}_d(F)|/|F^\times|$, we conclude in particular that

$$|F|^{d^2-1} \ll |G| \ll |F|^{d^2-1} \tag{4.1}$$

(this also follows from the Lang–Weil estimate, Proposition A.3). If b is not regular semisimple, then its characteristic polynomial has a repeated root. This constrains b to an algebraic hypersurface of $\mathrm{SL}_d(F)$ of complexity $O(1)$. This hypersurface has dimension $d^2 - 2$, so, by the Schwarz–Zippel lemma (see Proposition A.2), we have that at most $O(|F|^{d^2-2})$ elements of G are not regular semisimple. This is only $O(|F|^{-1})$ of the elements of G, so to prove (3.1) it suffices as before to consider the contribution of the regular semisimple b.

If b is regular semisimple, then the centralizer $Z(b)$ of b consists of the F-points of a $d-1$-dimensional torus $\overline{Z(b)}$ in $\mathrm{SL}_d(\overline{F})$, of complexity $O(1)$, defined over F. By the Lang–Weil bound (Proposition A.3), we have $|F|^{d-1} \ll |Z(b)| \ll |F|^{d-1}$. Arguing as in the previous section, we thus have

$$\mu_{b,h}(y) \ll |F|^{-d^2-d+2}|\{(g,c) \in G \times Z(b) : gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1} = y\}|. \tag{4.2}$$

Let $\phi_{b,h} : \mathrm{SL}_d(\overline{F}) \times \overline{Z(b)} \to \mathrm{SL}_d(\overline{F})$ be the map

$$\phi_{b,h}(g,c) := gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1}. \tag{4.3}$$

This is a regular map of complexity $O(1)$ from the $d^2 + d - 2$-dimensional irreducible variety $\mathrm{SL}_d(\overline{F}) \times \overline{Z(b)}$ to the $d^2 - 1$-dimensional variety $\mathrm{SL}_d(\overline{F})$.

Suppose that (b,h) is such that the map $\phi_{b,h}$ is dominant. Applying Proposition A.5, we see that there exists a subset Σ of $\mathrm{SL}_d(\overline{F}) \times \overline{Z(b)}$ which can be covered by $O(1)$ varieties of complexity $O(1)$ and dimension at most $d^2 + d - 3$, such that, for each $y \in \mathrm{SL}_d(\overline{F})$, the set

$$|\{(g,c) \in (\mathrm{SL}_d(\overline{F}) \times \overline{Z(b)})\backslash\Sigma : \phi_{b,h}(g,c) = y\}$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

is covered by $O(1)$ varieties of complexity $O(1)$ and dimension at most $d-1$. Applying the Schwarz–Zippel bound (Proposition A.2), we conclude that

$$|\{(g,c) \in (\mathrm{SL}_d(F) \times Z(b)) \backslash \Sigma : \phi_{b,h}(g,c) = y\}| \ll |F|^{d-1}$$

for all $y \in G$, and thus by (4.2) one has

$$\mu_{b,h}(y) \ll |F|^{-d^2+1} + |F|^{-d^2-d+2} \times |\{(g,c) \in (G \times Z(b)) \cap \Sigma : gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1} = y\}|.$$

By (4.1), we conclude (for C_0 large enough) that

$$\begin{aligned}\sum_{y \in G : \mu_{b,h}(y) \geq C_0/|G|} & \mu_{b,h}(y) \\ &\ll |F|^{-d^2-d+2} \sum_{y \in G} |\{(g,c) \in (G \times Z(b)) \cap \Sigma : gc^{-1}h^{-1}g^{-1}c^{-1}h^{-1} = y\}| \\ &= |F|^{-d^2-d+2} |(G \times Z(b)) \cap \Sigma|,\end{aligned}$$

and hence, by another application of the Schwarz–Zippel bound, we have

$$\sum_{y \in G : \mu_{b,h}(y) \geq C_0/|G|} \mu_{b,h}(y) \ll |F|^{-1}$$

when $\phi_{b,h}$ is dominant. On the other hand, when $\phi_{b,h}$ is not dominant, we may crudely bound

$$\sum_{y \in G : \mu_{b,h}(y) \geq C_0/|G|} \mu_{b,h}(y) \leq \sum_{y \in G} \mu_{b,h}(y) = 1.$$

To establish (3.1), it thus suffices to show that there are at most $O(|F|^{-1}|G|^2)$ pairs $(b,h) \in G \times G$ with b regular semisimple and $\phi_{b,h}$ not dominant.

Fix b to be a regular semisimple element. It suffices to show that $\phi_{b,h}$ is dominant for all but at most $O(|F|^{-1}|G|)$ values of $h \in G$; by the Schwarz–Zippel bound (Proposition A.2), it suffices to show that $\phi_{b,h}$ is dominant for all $h \in \mathrm{SL}_d(\overline{F})$ lying outside of $O(1)$ algebraic varieties of positive codimension and complexity $O(1)$. As this assertion only involves $\overline{F}$ and not F, we may now diagonalize b over $\overline{F}$, and work in a basis in which b is diagonal (with coefficients in $\overline{F}$ rather than in F). The torus $\overline{Z(b)}$ is now the group $T(\overline{F})$ of diagonal matrices in $\mathrm{SL}_d(\overline{F})$. It now suffices to establish the following claim.

PROPOSITION 4.1 (Quantitative generic nondegeneracy). *Let k be an algebraically closed field, and let $d \geq 1$; we allow all implied constants to depend on d. Then, for all $h \in \mathrm{SL}_d(k)$ outside of $O(1)$ algebraic varieties of positive codimension and complexity $O(1)$, the map $\tilde{\phi}_h : \mathrm{SL}_d(k) \times T(k) \to \mathrm{SL}_d(k)$*

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

defined by

$$\tilde{\phi}_h(g, c) := gc^{-1}h^{-1}g^{-1}c^{-1}h \tag{4.4}$$

is dominant, where $T(k)$ *denotes the group of diagonal matrices in* $\mathrm{SL}_d(k)$.

Indeed, by setting k equal to the algebraic closure $\overline{F}$ of F, and noting that $\phi_{b,h} = \tilde{\phi}_h h^{-2}$, the claim follows. (We have shifted $\tilde{\phi}_h$ in order to map the identity $(1, 1)$ to the identity 1.)

It turns out that, by using an ultraproduct argument, one can show that Proposition 4.1 is implied by the following, seemingly weaker, qualitative variant of that proposition, in which the uniform bounds on the exceptional set are dropped.

PROPOSITION 4.2 (Qualitative generic nondegeneracy). *Let* k *be an algebraically closed field, and let* $d \geq 1$. *Then, for generic* $h \in \mathrm{SL}_d(k)$ *(that is, for all* h *outside of a finite union of varieties of positive codimension), the map* $\tilde{\phi}_h \colon \mathrm{SL}_d(k) \times T(k) \to \mathrm{SL}_d(k)$ *defined by (4.4) is dominant.*

Indeed, if Proposition 4.1 failed, then one could find $d \geq 1$ and a sequence k_n of algebraically closed fields such that the set of $h \in \mathrm{SL}_d(k_n)$ for which $\tilde{\phi}_h$ fails to be dominant cannot be covered by n algebraic varieties of positive codimension and complexity at most n. Performing an ultraproduct with respect to a nonprincipal ultrafilter on the natural numbers (see [7, Appendix A]), we then obtain a new (and much larger) algebraically closed field k, with the property that the set of $h \in \mathrm{SL}_d(k)$ for which $\tilde{\phi}_h$ fails to be dominant cannot be covered by any finite number of algebraic varieties of positive codimension, contradicting Proposition 4.2. (Here, we use the continuity of irreducibility and dominance with respect to ultraproducts; see [7, Lemma A.2] and [7, Lemma A.7].)

It remains to prove Proposition 4.2. By the irreducibility of $\mathrm{SL}_d(\overline{F})$, it suffices to show that the derivative map

$$D\tilde{\phi}_h(1, 1) \colon \mathfrak{sl}_d(k) \times \mathfrak{t}(k) \to \mathfrak{sl}_d(k)$$

is full rank for generic $h \in \mathrm{SL}_d(k)$, where $\mathfrak{sl}_d(k)$ is the vector space of trace zero $d \times d$ matrices over k, and $\mathfrak{t}(k)$ is the subspace of $\mathfrak{sl}_d(k)$ consisting of diagonal matrices over k of trace zero. From the product rule and (4.4), we may evaluate $D\tilde{\phi}_h(1, 1)$ explicitly as

$$D\tilde{\phi}_h(1, 1)(X, Y) = X - h^{-1}Xh - Y - h^{-1}Yh$$

for $X \in \mathfrak{sl}_d(k)$ and $Y \in \mathfrak{t}(k)$.

We may restrict attention to those h which are regular semisimple (or, equivalently, those h whose characteristic polynomial has no repeated roots),

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

as the complement of this set is certainly contained in a finite number of algebraic varieties of positive codimension. We may thus diagonalize $h = ADA^{-1}$ for some $A \in \mathrm{SL}_d(k)$ and diagonal D with distinct diagonal entries. Then we have

$$D\tilde{\phi}_h(1,1)(X,Y) = A(X' - D^{-1}X'D - Y' - D^{-1}Y'D)A^{-1},$$

where $X' := A^{-1}XA$ and $Y' := A^{-1}YA$. We thus see that $D\tilde{\phi}_h(1,1)$ is full rank if and only if the map

$$(X', Y') \mapsto X' - D^{-1}X'D - Y' - D^{-1}Y'D$$

is a full rank map from $\mathfrak{sl}_d(\overline{F}) \times A^{-1}\mathfrak{t}(\overline{F})A$ to $\mathfrak{sl}_d(\overline{F})$. It thus suffices to show that this map is full rank for generic $A \in \mathrm{SL}_d(k)$ and $D \in T(k)$.

As D is a diagonal matrix with distinct diagonal entries, we see that the image of $\mathfrak{sl}_d(k)$ under the map $X' \mapsto X' - D^{-1}X'D$ is the space of all matrices that vanish on the diagonal. To show that $D\tilde{\phi}_h(1,1)$ has full rank, it thus suffices to show that the map $Y' \mapsto \mathrm{diag}(Y' + D^{-1}Y'D)$ has full rank from $A^{-1}\mathfrak{t}(\overline{F})A$ to $\mathfrak{t}(\overline{F})$. Since $\mathrm{diag}(Y' + D^{-1}Y'D) = 2\,\mathrm{diag}(Y')$, it suffices to show that the diagonal map $Y' \mapsto \mathrm{diag}(Y')$ has full rank from $A^{-1}\mathfrak{t}(\overline{F})A$ to $\mathfrak{t}(\overline{F})$ for generic $A \in \mathrm{SL}_d(k)$. As this is clearly a Zariski-open algebraic constraint, and contains the case $A = 1$, we conclude that one has full rank for generic A, and the claim follows.

5. Expansion

In the remarkable paper of Bourgain and Gamburd [6], the quasirandomness properties of $\mathrm{SL}_2(F)$, combined with the product theory in such groups (see [14]), were used to establish spectral gaps for the generators of various Cayley graphs. In our notation, the results of [6] established spectral gap results, a typical one of which is the assertion that, with probability $1 - o_{p\to\infty}(1)$, one has

$$\left\| \frac{1}{4}(\delta_a + \delta_b + \delta_{a^{-1}} + \delta_{b^{-1}}) \right\|_{S(\mathrm{SL}_2(F_p))} \leq 1 - c$$

for some absolute constant $c > 0$, where F_p is a finite field of prime order and a, b are chosen uniformly at random from $\mathrm{SL}_2(F_p)$. This result has since been generalized in a number of different directions; see [18] for a survey.

In this section, we establish some related expansion results, but, instead of a probability measure (such as $\frac{1}{4}(\delta_a + \delta_b + \delta_{a^{-1}} + \delta_{b^{-1}})$) supported on a small number of points, we will establish spectral bounds on (quasi)probability measures distributed more or less uniformly on subvarieties V of SL_d; this will play an important role in the proof of Theorem 1.5 in later sections. The main result is that, as long as V is not 'trapped' in an algebraic subgroup of SL_d

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

(or a coset thereof), there is a spectral norm bound which gains a power of $|F|$ over the trivial bound. The arguments are very much in the spirit of Bourgain and Gamburd [6], with the main ingredients being 'escape from subvarieties', quasirandomness, and some basic algebraic geometry. However, due to the algebraic structure of the measures being studied, combinatorial tools such as the product theorem of Helfgott [14] are not required in this argument (though they could certainly be deployed in order to prove more general results, in which the measure in question is not assumed to be adapted to an algebraic subvariety).

More precisely, we will establish the following result.

PROPOSITION 5.1 (Expansion from subvarieties). *Let k be an algebraically closed field, and let F be a finite subfield of k. Let $V \subset \mathrm{SL}_d(k)$ be an irreducible algebraic variety defined over k of complexity at most M. Suppose that V is not contained in any coset Hg of a proper algebraic subgroup H of $\mathrm{SL}_d(k)$. Then one has*

$$\|\mu\|_{S(\mathrm{SL}_d(F))} \ll_{d,M} |F|^{\dim(V)-c} \|\mu\|_{L^\infty(V \cap \mathrm{SL}_d(F))}$$

for all $\mu: \mathrm{SL}_d(F) \to \mathbf{C}$ supported on $V \cap \mathrm{SL}_d(F)$, where $c > 0$ depends only on d.

Recall that $\|\|_{S(G)}$ is the reduced spectral norm, defined in (2.1).

Proof. We perform a downward induction on $\dim(V)$, which is an integer between 0 and $\dim(\mathrm{SL}_d) = d^2 - 1$. When $\dim(V) = \dim(\mathrm{SL}_d)$, the claim follows from (2.4), (4.1), and Proposition 1.2. Now suppose that $\dim(V) < \dim(\mathrm{SL}_d)$, and that the claim has already been proven for all larger values of $\dim(V)$.

We normalize $\|\mu\|_{L^\infty(V \cap \mathrm{SL}_d(F))} := |F|^{-\dim(V)}$, and allow all implied constants to depend on d and M, so our task is now to show that

$$\|\mu\|_{S(\mathrm{SL}_d(F))} \ll |F|^{-c}.$$

Recall the *TT^* identity*

$$\|TT^*\|_{\mathrm{op}} = \|T\|_{\mathrm{op}}^2$$

whenever T is a bounded linear operator between Hilbert spaces. Applying this to the convolution operator $f \mapsto f * \mu$ on the Hilbert space of mean zero functions on $L^2(G)$, we conclude that

$$\|\mu * \tilde{\mu}\|_{S(\mathrm{SL}_d(F))} = \|\mu\|_{S(\mathrm{SL}_d(F))}^2,$$

where $\tilde{\mu}: G \to \mathbf{C}$ is the function $\tilde{\mu}(g) := \overline{\mu(g^{-1})}$. It will thus suffice to show that

$$\|\mu * \tilde{\mu}\|_{S(\mathrm{SL}_d(F))} \ll |F|^{-c}$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

for some $c > 0$ depending only on m, d. (Note that, as there are only $O(1)$ different values of $\dim(V)$, we may allow the value of the constant c to change with each step of the induction.)

We consider the product map $\phi: V \times V \to \mathrm{SL}_d(k)$ given by $\phi(v, w) := vw^{-1}$, and let W' be the Zariski closure of $\phi(V \times V)$. As $V \times V$ is irreducible, W' is also irreducible. As W' contains a translate of V, we have $\dim(W') \geq \dim(V)$. We claim that we in fact have strict inequality $\dim(W') > \dim(V)$. To see this, suppose for contradiction that $\dim(W') = \dim(V)$. Then, for each $w \in V$, Vw^{-1} is contained in the irreducible variety W', and has the same dimension as W', and so $Vw^{-1} = W'$ for all $w \in V$. This implies that $W'(W')^{-1} = \phi(V \times V) \subset W'$, or in other words that W' forms a group, and is thus a proper algebraic subgroup of $\mathrm{SL}_d(k)$. But V is contained in a coset of W, contradicting the hypothesis on V. Thus we have $\dim(W') > \dim(V)$.

We now apply Proposition A.5, to conclude that W' has complexity $O(1)$, and that there is a subset Σ of $V \times V$ covered by $O(1)$ varieties of complexity $O(1)$ and dimension strictly less than $2\dim(V)$, such that, for each $w \in W'$, the set $\{(v, v') \in V \times V \backslash \Sigma : \phi(v, v') = w\}$ is contained in $O(1)$ varieties of complexity $O(1)$ and dimension at most $2\dim(V) - \dim(W')$. Applying the Schwarz–Zippel bound (Proposition A.2), we conclude that

$$|\Sigma \cap (G \times G)| \ll |F|^{2\dim(V)-1} \tag{5.1}$$

and

$$|\{(v, v') \in ((V \times V) \cap (G \times G)) \backslash \Sigma : \phi(v, v') = w\}| \ll |F|^{2\dim(V)-\dim(W')}. \tag{5.2}$$

Next, we expand

$$\mu * \tilde{\mu}(w) = \sum_{(v,v') \in (V \times V) \cap (G \times G) : \phi(v,v') = w} \mu(v)\overline{\mu(v')},$$

and then decompose

$$\mu * \tilde{\mu} = \mu_1 + \mu_2,$$

where

$$\mu_1(w) := \sum_{(v,v') \in \Sigma \cap (G \times G) : \phi(v,v') = w} \mu(v)\overline{\mu(v')}$$

and

$$\mu_2(w) := \sum_{(v,v') \in ((V \times V) \cap (G \times G)) \backslash \Sigma : \phi(v,v') = w} \mu(v)\overline{\mu(v')}.$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

As $\|\mu\|_{L^\infty(V)} = |F|^{-\dim(V)}$, we see that

$$\begin{aligned}\|\mu_1\|_{\ell_1(G)} &\leq \sum_{(v,v')\in\Sigma\cap(G\times G)} |F|^{-\dim(V)}|F|^{-\dim(V)} \\ &\ll |F|^{-1},\end{aligned} \tag{5.3}$$

thanks to (5.1). By (2.3), we thus have

$$\|\mu_1\|_{S(G)} \ll |F|^{-1}.$$

Next, from (5.2) and the normalization $\|\mu\|_{L^\infty(V)} = |F|^{-\dim(V)}$, we have

$$\mu_2(w) \ll |F|^{2\dim(V)-\dim(W')}|F|^{-\dim(V)}|F|^{-\dim(V)} = |F|^{-\dim(W')}$$

for all $w \in G$. As μ_2 is supported on W', we conclude from the induction hypothesis that

$$\|\mu_2\|_{S(G)} \ll |F|^{-c}$$

for some $c > 0$ depending only on d, and the claim follows. (Note that, as W' contains a translate of V, it cannot itself be contained in a coset of a proper algebraic subgroup of G.) □

We remark that the above proof in fact allows one to take $c := 2^{-2^{\dim(V)-d}}$.

We will apply Proposition 5.1 in the case of a function μ supported on a conjugacy class, as in the following corollary.

COROLLARY 5.2. *Let F be a finite field, let $d \geq 2$, and let $a \in \mathrm{SL}_d(F)$ be noncentral (that is, a is not a multiple of the identity). Let $C(a) := \{gag^{-1} : g \in \mathrm{SL}_d(F)\}$ be the conjugacy class of a. Then*

$$\|1_{C(a)}\|_{S(\mathrm{SL}_d(F))} \ll_d |F|^{-c}|C(a)|$$

for some $c > 0$ depending only on d.

Proof. We allow all implied constants to depend on d. We apply Proposition 5.1 with k equal to the algebraic closure of F, and V equal to the closed conjugacy class $\overline{C(a)} := \overline{\{gag^{-1} : g \in \mathrm{SL}_d(k)\}}$. It is clear that V is an irreducible algebraic variety defined over k of complexity $O(1)$; the irreducibility follows since $\mathrm{SL}_d(k)$ is irreducible and the map $g \mapsto gag^{-1}$ is algebraic. Proposition 5.1 will give the desired claim unless $\overline{C(a)}$ is contained in a coset Hg of a proper algebraic subgroup H of $\mathrm{SL}_d(k)$. But this implies that H contains $\overline{C(a)} \cdot \overline{C(a)}^{-1}$, which implies that the group N generated by $\overline{C(a)} \cdot \overline{C(a)}^{-1}$ is a proper subgroup of $\mathrm{SL}_d(k)$. But this group is conjugation invariant, and thus normal. It is a classical fact (see for example [15]) that the algebraic group $\mathrm{SL}_d(k)$ is almost simple, in the sense that the only normal subgroups are finite (in fact, the maximal normal

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

subgroup is the center, or equivalently the quotient $PSL_d(k)$ is simple). This implies that $\overline{C(a)}$ is finite. But this contradicts the hypothesis that a is not central, and the claim follows. □

REMARK 5.3. A standard application of Schur's lemma gives the identity

$$\mathbf{E}_{b\in C(a)}\rho(b) = \frac{1}{\dim(V)}(\operatorname{trace}\rho(a))I_V$$

for any nontrivial irreducible unitary representation $\rho\colon \mathrm{SL}_d(F) \to U(V)$, where I_V denotes the identity operator on V. From this and Remark 2.1 we see that Corollary 5.2 is equivalent to the assertion that $|\operatorname{trace}\rho(a)| \ll_d |F|^{-c}\dim(V)$ for any nontrivial irreducible representation $\rho\colon \mathrm{SL}_d(F) \to U(V)$ and any noncentral a. It is likely that this result could also be established directly (with an optimal value of c) from the representation theory of $\mathrm{SL}_d(F)$, but we will not do so here.

6. A reduction to a Borel group

We will abbreviate $o_{|F|\to\infty}()$ as $o()$ throughout the rest of this paper.

We now begin the proof of Theorem 1.5 by making some reductions. The first is to use the Cauchy–Schwarz inequality to reduce Theorem 1.5 to a seemingly weaker statement in which the absolute values have been moved outside of the g averaging. In other words, we will deduce Theorem 1.5 from the following statement.

THEOREM 6.1. *Let F be a finite field, and set $G := \mathrm{SL}_2(F)$. Let S denote the set of all elements of $\mathrm{SL}_2(F)$ that are diagonalizable over F. Then, for any functions $f_0, f_1, f_2, f_3\colon G \to \mathbf{C}$, we have*

$$\left|\mathbf{E}_{g\in S}\mathbf{E}_{x\in G}\prod_{i=0}^{3} f_i(xg^{i-1}) - \prod_{i=0}^{3}\mathbf{E}_G f_i\right| \ll o\left(\prod_{i=0}^{3}\|f_i\|_{L^\infty(G)}\right).$$

Let us assume Theorem 6.1 for now, and see how it implies Theorem 1.5. If f_3 is constant, then the claim follows from Theorem 1.4, so we may assume without loss of generality that f_3 has mean zero. We may take the f_i to be real valued, and also normalize $\|f_i\|_{L^\infty(G)} = 1$ for each i. Our task is now to show that

$$\mathbf{E}_{g\in S}\left|\mathbf{E}_{x\in G}\prod_{i=0}^{3} f_i(xg^{i-1})\right| = o(1).$$

By the Cauchy–Schwarz inequality, it suffices to show that

$$\mathbf{E}_{g\in S}\left|\mathbf{E}_{x\in G}\prod_{i=0}^{3} f_i(xg^{i-1})\right|^2 = o(1),$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

which we square as

$$\mathbf{E}_{g\in S}\mathbf{E}_{x,y\in G}\prod_{i=0}^{3}f_i(xg^{i-1})f_i(yg^{i-1}) = o(1).$$

Substituting $y = hx$, we can rewrite the left-hand side as

$$\mathbf{E}_{h\in G}\mathbf{E}_{g\in S}\mathbf{E}_{x\in G}\prod_{i=0}^{3}f_i(xg^{i-1})f_i(hxg^{i-1}).$$

Applying Theorem 6.1, we have

$$\mathbf{E}_{g\in S}\mathbf{E}_{x\in G}\prod_{i=0}^{3}f_i(xg^{i-1})f_i(hxg^{i-1}) = \prod_{i=0}^{3}\mathbf{E}_{x\in G}f_i(x)f_i(hx) + o(1)$$

for each $h \in G$, so it suffices to show that

$$\left|\mathbf{E}_{h\in G}\prod_{i=0}^{3}\mathbf{E}_{x\in G}f_i(x)f_i(hx)\right| = o(1).$$

We can bound the left-hand side in magnitude by

$$\mathbf{E}_{h\in G}|\mathbf{E}_{x\in G}f_3(x)f_3(hx)|,$$

and the claim now follows from Lemma 1.3 (applied to the reversed function $x \mapsto f_3(x^{-1})$).

It remains to establish Theorem 6.1. We will deduce it from the following variant theorem on the standard Borel subgroup B of $\mathrm{SL}_d(F)$.

THEOREM 6.2. *Let F be a finite field, and let B be the subgroup of matrices in $\mathrm{SL}_2(F)$ which are upper triangular. Let U be the normal subgroup of B consisting of matrices which are equal to the identity matrix except possibly at the upper right entry. Let $f_0, \dots, f_3 : B \to \mathbf{C}$. Then*

$$\Lambda_{4,B}(f_0, \dots, f_3) = \Lambda_{4,B}(f_0 * \mu_U, \dots, f_3 * \mu_U) + o(\|f_0\|_{L^\infty(B)} \dots \|f_3\|_{L^\infty(B)}),$$

where $\mu_U := (1/|U|)1_U$.

Let us assume Theorem 6.2 for now, and show how it implies Theorem 6.1. We may again assume that f_3 has mean zero, and that the f_i are real valued with $\|f_i\|_{L^\infty(G)} = 1$ for each i. Our task is to show that

$$\left|\mathbf{E}_{g\in S}\mathbf{E}_{x\in G}\prod_{i=0}^{3}f_i(xg^{i-1})\right| = o(1).$$

The first task is to replace the set S by the set B as follows. Observe that B is the space of all matrices in $\mathrm{SL}_2(F)$ that fix the span $\mathrm{span}(e_2)$ of the second vector e_2

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

of the standard basis e_1, e_2 of F^2. Any conjugate gBg^{-1} of B, where $g \in \mathrm{SL}_2(F)$, would fix another line; this new line would be identical to the original line $\mathrm{span}(e_2)$ precisely when $g \in B$, so the total number of such conjugates is

$$|\mathrm{SL}_2(F)|/|B| = (1 + O(|F|^{-1}))|F|.$$

If $g \in S$ is regular semisimple, then it has two distinct one-dimensional eigenspaces in F, and thus preserves $2! = 2$ distinct lines. As such, it lies in gBg^{-1} for $2|B|$ different values of B. We thus see that the number of regular semisimple elements of S is equal to $|G|/2|B|$ times the number of regular semisimple elements of B. An element of B is regular semisimple if and only if its diagonal entries are distinct, so we see that the proportion of elements of B that are regular semisimple is $1 - O(|F|^{-1})$. We conclude that there are $(\frac{1}{2} + O(|F|^{-1}))|G|$ regular semisimple elements of S. As all but $O(|F|^{-1}|G|)$ elements of G (and hence of S) are regular semisimple, we thus see that

$$\mathbf{E}_{g \in S} f(g) = \mathbf{E}_{g \in G} \mathbf{E}_{h \in gBg^{-1}} f(h) + O(|F|^{-1})$$

for any function $f\colon G \to \mathbf{C}$ of magnitude $O(1)$. It will thus suffice to show that

$$\mathbf{E}_{g \in G} \mathbf{E}_{h \in gBg^{-1}} \mathbf{E}_{x \in G} \prod_{i=0}^{3} f_i(xh^{i-1}) = o(1).$$

Fix $g \in G$. By foliating G into left cosets $agBg^{-1}$ of gBg^{-1}, and applying Theorem 6.2 (conjugated by g) to each coset, we see that

$$\mathbf{E}_{h \in gBg^{-1}} \mathbf{E}_{x \in agBg^{-1}} \prod_{i=0}^{3} f_i(xh^{i-1}) = \mathbf{E}_{h \in gBg^{-1}} \mathbf{E}_{x \in agBg^{-1}} \prod_{i=0}^{3} (f_i * \mu_{gUg^{-1}})(xh^{i-1}) + o(1)$$

for each a. It thus suffices to show that

$$\mathbf{E}_{g \in G} \mathbf{E}_{h \in gBg^{-1}} \mathbf{E}_{x \in G} \prod_{i=0}^{3} (f_i * \mu_{gUg^{-1}})(xh^{i-1}) = o(1).$$

Applying the crude bound

$$\left| \mathbf{E}_{h \in gBg^{-1}} \mathbf{E}_{x \in G} \prod_{i=0}^{3} (f_i * \mu_{gUg^{-1}})(xh^{i-1}) \right| \leq \mathbf{E}_{x \in G} |f_3 * \mu_{gUg^{-1}}(x)|,$$

it suffices to show that

$$\mathbf{E}_{g \in G} \mathbf{E}_{x \in G} |f_3 * \mu_{gUg^{-1}}(x)| = o(1).$$

By the Cauchy–Schwarz inequality, it suffices to show that

$$\mathbf{E}_{g \in G} \mathbf{E}_{x \in G} |f_3 * \mu_{gUg^{-1}}(x)|^2 = o(1).$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

From the identity

$$\mathbf{E}_{x\in G}|f_3 * \mu_{gUg^{-1}}(x)|^2 = \mathbf{E}_{x\in G} f_3(x)(f_3 * \mu_{gUg^{-1}})(x),$$

it suffices to show that

$$|\mathbf{E}_{g\in G}\mathbf{E}_{x\in G} f_3(x)(f_3 * \mu_{gUg^{-1}})(x)| = o(1).$$

By definition of the reduced spectral norm, the left-hand side is bounded by

$$\|\mathbf{E}_{g\in G}\mu_{gUg^{-1}}\|_S.$$

Observe that

$$\mathbf{E}_{g\in G}\mu_{gUg^{-1}} = \mathbf{E}_{u\in U}\mathbf{E}_{g\in G}\delta_{gug^{-1}} = \mathbf{E}_{u\in U}\frac{1}{|C(u)|}1_{C(u)},$$

and so, by Minkowski's inequality,

$$\|\mathbf{E}_{g\in G}\mu_{gUg^{-1}}\|_S \leq \mathbf{E}_{u\in U}\frac{1}{|C(u)|}\|1_{C(u)}\|_S.$$

By Corollary 5.2, we may bound $(1/|C(u)|)\|1_{C(u)}\|_S$ by $|F|^{-c}$ for some $c > 0$ depending only on d, except when u is the identity element, in which case we have the trivial bound of 1. As U has cardinality $|F|$, we obtain a net bound of $O(|F|^{-1} + |F|^{-c})$, and the claim follows.

It remains to establish Theorem 6.2. This is the purpose of the remaining sections of the paper.

7. Progressions in a Borel group

We now prove Theorem 6.2.

By splitting each function f_i into functions that are constant along cosets of U, or have mean zero along cosets of U, we see that it suffices to show that

$$\Lambda_{4,B}(f_0, \ldots, f_3) = o(\|f_0\|_{L^\infty(B)} \ldots \|f_3\|_{L^\infty(B)})$$

whenever at least one of f_0, f_1, f_2, f_3 has mean zero along cosets of U. By the symmetry

$$\Lambda_{4,B}(f_0, \ldots, f_3) = \Lambda_{4,B}(f_3, \ldots, f_0),$$

we may assume that f_{i_0} has mean zero along cosets of U for some $i_0 \in \{2, 3\}$. We may also take f_0, f_1, f_2, f_3 to be real valued with $L^\infty(B)$ norm of 1, so our task is to show that

$$\mathbf{E}_{x,g\in B} f_0(x) f_1(xg) f_2(xg^2) f_3(xg^3) = o(1).$$

We will take advantage of the short exact sequence

$$0 \to F \to B \to F^\times \to 0$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

between the additive group $F = (F, +)$, the Borel group B, and the multiplicative group $F^\times := (F\backslash\{0\}, \cdot)$, given by the inclusion map $\psi: F \to B$ and the projection map $\pi: B \to F^\times$ defined by the formulae

$$\psi(a) := \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}$$

and

$$\pi\left(\begin{pmatrix} t & a \\ 0 & t^{-1} \end{pmatrix}\right) = t^{-1}.$$

For any $a, b \in F$, we can make the change of variables $(x, g) \mapsto (\psi(a)x, \psi(b)g)$, and write

$$\begin{aligned}\mathbf{E}_{x,g\in B} f_0(x) f_1(xg) f_2(xg^2) f_3(xg^3) = \mathbf{E}_{x,g\in B} & f_0(\psi(a)x) f_1(\psi(a)x\psi(b)g) \\ & \times f_2(\psi(a)x\psi(b)g\psi(b)g) \\ & \times f_3(\psi(a)x\psi(b)g\psi(b)g\psi(b)g).\end{aligned}$$

By using the identity

$$x\psi(b) = \psi(\pi(x)^2 b)x$$

for any $x \in B$ and $b \in F$, we can rewrite the above identity as

$$\begin{aligned}&\mathbf{E}_{x,g\in B} f_0(x) f_1(xg) f_2(xg^2) f_3(xg^3) \\ &\quad = \mathbf{E}_{x,g\in B} f_0(\psi(a)x) f_1(\psi(a + \pi(x)^2 b)xg) f_2(\psi(a + \pi(x)^2 b + \pi(xg)^2 b)xg^2) \\ &\qquad \times f_3(\psi(a + \pi(x)^2 b + \pi(xg)^2 b + \pi(xg^2)b)xg^3).\end{aligned}$$

On averaging in a, b, we conclude that

$$\begin{aligned}\mathbf{E}_{x,g\in B} f_0(x) f_1(xg) f_2(xg^2) f_3(xg^3) = \mathbf{E}_{x,g\in B} \mathbf{E}_{a,b\in F} & f_{0,x}(a) f_{1,xg}(a + \pi(x)^2 b) \\ & \times f_{2,xg^2}(a + \pi(x)^2 b + \pi(xg)^2 b) \\ & \times f_{3,xg^3}(a + \pi(x)^2 b + \pi(xg)^2 b + \pi(xg^2)^2 b),\end{aligned}$$

where $f_{i,x}: F \to \mathbf{R}$ are the functions

$$f_{i,x}(a) := f_i(\psi(a)x).$$

By dilating b by $\pi(x)^2$, we may simplify the above expression slightly as

$$\begin{aligned}&\mathbf{E}_{x,g\in B} \mathbf{E}_{a,b\in F} f_{0,x}(a) f_{1,xg}(a + b) \\ &\quad \times f_{2,xg^2}(a + (1 + \pi(g)^2)b) f_{3,xg^3}(a + (1 + \pi(g)^2 + \pi(g)^4)b).\end{aligned}$$

As is well known, the inner average has too high a 'complexity' to be directly treated by Fourier analysis. However, following Gowers [9], we may reduce it to

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

a form that is tractable to Fourier analysis after applying the Cauchy–Schwarz inequality. Indeed, from that inequality we can bound the preceding expression in magnitude by

$$(\mathbf{E}_{x,g\in B}\mathbf{E}_{a\in F}|\mathbf{E}_{b\in F}f_{1,xg}(a+b)f_{2,xg^2}(a+(1+\pi(g)^2)b) \times f_{3,xg^3}(a+(1+\pi(g)^2+\pi(g)^4)b)|^2)^{1/2}.$$

We may expand this expression as

$$(\mathbf{E}_{x,g\in B}\mathbf{E}_{a,b,b'\in F}f_{1,xg}(a+b)f_{1,xg}(a+b') \times f_{2,xg^2}(a+(1+\pi(g)^2)b)f_{2,xg^2}(a+(1+\pi(g)^2)b') \times f_{3,xg^3}(a+(1+\pi(g)^2+\pi(g)^4)b)f_{3,xg^3}(a+(1+\pi(g)^2+\pi(g)^4)b'))^{1/2}.$$

Writing $b'=b+h$ and shifting x by g, this becomes

$$(\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}\mathbf{E}_{a,b\in F}\Delta_h f_{1,x}(a+b)\Delta_{(1+\pi(g)^2)h}f_{2,xg}(a+(1+\pi(g)^2)b) \times\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2}(a+(1+\pi(g)^2+\pi(g)^4)b))^{1/2},$$

where $\Delta_h f(a):=f(a)f(a+h)$.

Shifting a by b, then dilating b by $\pi(g)^{-2}$, we may simplify this slightly as

$$(\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}\mathbf{E}_{a,b\in F}\Delta_h f_{1,x}(a)\Delta_{(1+\pi(g)^2)h}f_{2,xg}(a+b) \times\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2}(a+(1+\pi(g)^2)b))^{1/2},$$

and so our task is now to show that

$$\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}\mathbf{E}_{a,b\in F}\Delta_h f_{1,x}(a)\Delta_{(1+\pi(g)^2)h}f_{2,xg}(a+b) \times\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2}(a+(1+\pi(g)^2)b)=o(1). \tag{7.1}$$

The next step is Fourier expansion. Consider the trilinear form

$$\mathbf{E}_{a,b\in F}H_1(a)H_2(a+b)H_3(a+(1+\pi(g)^2)b)$$

for some functions $H_1,H_2,H_3:F\to\mathbf{C}$. Using some arbitrary nondegenerate bilinear form $\cdot:F\times F\to\mathbf{R}/\mathbf{Z}$, we can form the Fourier series

$$H_i(a)=\sum_{\xi\in F}\hat{H}_i(\xi)e(\xi\cdot a)$$

for $i=1,2,3$, where $e(x):=e^{2\pi ix}$ and

$$\hat{H}_i(\xi)=\mathbf{E}_{a\in F}H_i(a)e(-\xi\cdot a).$$

Inserting these Fourier series and simplifying, we arrive at the identity

$$\mathbf{E}_{a,b\in F}H_1(a)H_2(a+b)H_3(a+(1+\pi(g)^2)b) =\sum_{\xi\in F}\hat{H}_1(\xi)\hat{H}_2(-(1+\pi(g)^{-2})\xi)\hat{H}_3(\pi(g)^{-2}\xi).$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

We may thus write the left-hand side of (7.1) as

$$\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}\sum_{\xi\in F}(\Delta_h f_{1,x})^\wedge(\xi)(\Delta_{(1+\pi(g)^2)h}f_{2,xg})^\wedge(-(1+\pi(g)^{-2})\xi)$$
$$\times(\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2})^\wedge(\pi(g)^{-2}\xi).$$

Splitting off the $\xi=0$ and $\xi\neq 0$ terms, we see that, to prove (7.1), it will suffice to establish the bounds

$$\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}(\Delta_h f_{1,x})^\wedge(0)(\Delta_{(1+\pi(g)^2)h}f_{2,xg})^\wedge(0)(\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2})^\wedge(0)=o(1) \tag{7.2}$$

and

$$\mathbf{E}_{x,g\in B}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times}|(\Delta_h f_{1,x})^\wedge(\xi)||(\Delta_{(1+\pi(g)^2)h}f_{2,xg})^\wedge(-(1+\pi(g)^{-2})\xi)|$$
$$\times|(\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{3,xg^2})^\wedge(\pi(g)^{-2}\xi)|=o(1). \tag{7.3}$$

7.1. The contribution of the zero frequency. We now prove (7.2). We have

$$(\Delta_h f_{1,x})^\wedge(0)=\mathbf{E}_{a\in F}f_{1,x}(a)f_{1,x}(a+h),$$

and thus, by Fourier expansion,

$$(\Delta_h f_{1,x})^\wedge(0)=\sum_{\xi_1\in F}|\hat{f}_{1,x}(\xi_1)|^2e(\xi_1\cdot h).$$

Similarly we have

$$(\Delta_{(1+\pi(g)^2)h}f_{2,xg})^\wedge(0)=\sum_{\xi_2\in F}|\hat{f}_{2,xg}(\xi_2)|^2e((1+\pi(g)^2)^*\xi_2\cdot h)$$

and

$$(\Delta_{(1+\pi(g)^2+\pi(g)^4)h}f_{2,xg})^\wedge(0)=\sum_{\xi_3\in F}|\hat{f}_{3,xg^2}(\xi_3)|^2e((1+\pi(g)^2+\pi(g)^4)$$
$$+\rho(xg^2x^{-1})^*\xi_3\cdot h).$$

Inserting these identities and performing the h averaging, we conclude that the left-hand side of (7.2) can be rewritten as

$$\mathbf{E}_{x,g\in B}\sum_{\xi_1,\xi_2,\xi_3\in F:\xi_1+(1+\pi(g)^2)\xi_2+(1+\pi(g)^2+\pi(g)^4)\xi_3=0}|\hat{f}_{1,x}(\xi_1)|^2|\hat{f}_{2,xg}(\xi_2)|^2|\hat{f}_{3,xg^2}(\xi_3)|^2.$$

Recall that f_{i_0} was assumed to have mean zero on cosets of H, which implies that we may restrict ξ_{i_0} to be nonzero. We note that the quantity $|\hat{f}_{i,x}(\xi)|^2$ is unchanged

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

if one multiplies x on the left (or right) by an element of U, and so we may write

$$|\hat{f}_{i,x}(\xi)|^2 = \mu_{i,\pi(x)}(\xi)$$

for some nonnegative quantity $\mu_{i,t}(\xi)$, defined for $i = 1, 2, 3$, $t \in F^\times$, and $\xi \in F$. We can then simplify the previous expression as

$$\mathbf{E}_{s,t\in F^\times} \sum_{\xi_1,\xi_2,\xi_3\in F:\xi_1+(1+t^2)\xi_2+(1+t^2+t^4)\xi_3=0;\xi_{i_0}\neq 0} \mu_{1,s}(\xi_1)\mu_{2,st}(\xi_2)\mu_{3,st^2}(\xi_3). \quad (7.4)$$

To show that this expression is $o(1)$, it will suffice to establish the combinatorial bound

$$\mathbf{E}_{s,t\in F^\times} 1_{\eta_1(s)+(1+t^2)\eta_2(st)+(1+t^2+t^4)\eta_3(st^2)=0} = o(1) \quad (7.5)$$

for any choice of functions $\eta_i: F^\times \to F$ for $i = 1, 2, 3$, with η_{i_0} nonzero. Indeed, by the Plancherel identity, we have

$$\sum_\xi \mu_{i,s}(\xi) \leq 1$$

for all $i = 1, 2, 3$ and $s \in F^\times$, with $\mu_{i_0,s}(0) = 0$, so we may find random functions $\eta_i: F^\times \to F$ with η_{i_0} nowhere vanishing, and with the property that

$$\mu_{i,s}(\xi) \leq \mathbf{P}(\eta_i(s) = \xi)$$

for all $i = 1, 2, 3$ and $s \in F^\times$. Applying (7.5) with these functions, and taking expectations, we conclude that the quantity (7.4) is $o(1)$, as desired.

It remains to establish (7.5), which is a bound of 'sum-product' type, in that it is asserting a certain combinatorial incompatibility between the multiplicative and additive structures on F. Assume for contradiction that we can find arbitrarily large finite fields F and functions $\eta_1, \eta_2, \eta_3: F^\times \to F$ with η_{i_0} nowhere vanishing, for which

$$\mathbf{E}_{s,t\in F^\times} 1_{\eta_1(s)+(1+t^2)\eta_2(st)+(1+t^2+t^4)\eta_3(st^2)=0} \gg 1.$$

Fix $F, \eta_1, \eta_2, \eta_3$. Let $A \subset (F^\times)^2$ be the set of all pairs (s, t) for which

$$\eta_1(s) + (1 + t^2)\eta_2(st) + (1 + t^2 + t^4)\eta_3(st^2) = 0,$$

and thus $|A| \gg |F^\times|^2$. Applying the multidimensional Szemerédi theorem (Theorem B.1) to the multiplicative group $F^\times$, we conclude that there are $\gg |F|^3$ triples (s, t, r) with the property that $(sr^i, tr^j) \in A$ for all $-100 \leq i, j \leq 100$ (say), and thus

$$\eta_1(sr^i) + (1 + r^{2j}t^2)\eta_2(str^{i+j}) + (1 + r^{2j}t^2 + r^{4j}t^4)\eta_3(st^2r^{i+2j}) = 0 \quad (7.6)$$

for all $-100 \leq i, j \leq 100$. We will eliminate the η_i terms from (7.6) (taking advantage of the nonvanishing nature of η_{i_0}) to obtain a nontrivial algebraic

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

constraint on s, t, r, which will contradict the assertion that $\gg |F|^3$ triples (s, t, r) exist with this property if $|F|$ is large enough.

We turn to the details. Fix s, t, r obeying (7.6). If we abbreviate $\eta_k(st^{k-1}r^i)$ as $c_k(i)$, and also write $\alpha_j := 1 + r^{2j}t^2$ and $\beta_j := 1 + r^{2j}t^2 + r^{4j}t^4$, we have

$$c_1(i) + \alpha_j c_2(i+j) + \beta_j c_3(i+2j) = 0$$

for all $-100 \leq i, j \leq 100$. In particular, applying this identity for j and $j+1$ and subtracting, we have

$$\alpha_{j+1}c_2(i+j+1) - \alpha_j c_2(i+j) = \beta_j c_3(i+2j) - \beta_{j+1}c_3(i+2j+2)$$

for all $-90 \leq i, j \leq 90$ (say). Replacing (i, j) by $(i-2, j+2)$, $(i+2, j-1)$, and $(i, j+1)$, we obtain the system of four equations

$$\alpha_{j+1}c_2(i+j+1) - \alpha_j c_2(i+j) = \beta_j c_3(i+2j) - \beta_{j+1}c_3(i+2j+2) \quad (7.7)$$

$$\alpha_{j+3}c_2(i+j+1) - \alpha_{j+2}c_2(i+j) = \beta_{j+2}c_3(i+2j+2) - \beta_{j+3}c_3(i+2j+4) \quad (7.8)$$

$$\alpha_j c_2(i+j+2) - \alpha_{j-1}c_2(i+j+1) = \beta_{j-1}c_3(i+2j) - \beta_j c_3(i+2j+2) \quad (7.9)$$

$$\alpha_{j+2}c_2(i+j+2) - \alpha_{j+1}c_2(i+j+1) = \beta_{j+1}c_3(i+2j+2) - \beta_{j+2}c_3(i+2j+4) \quad (7.10)$$

for all $-80 \leq i, j \leq 80$ (say).

We now eliminate the various c_2 factors in this system to obtain a linear recurrence in the c_j. Multiplying (7.7) by α_{j+2} and (7.8) by α_j, and subtracting to eliminate the $c_2(i+j)$ term, we conclude that

$$\begin{aligned}(\alpha_{j+1}\alpha_{j+2} - \alpha_{j+3}\alpha_j)c_2(i+j+1) &= \beta_j\alpha_{j+2}c_3(i+2j) - (\beta_{j+1}\alpha_{j+2} + \beta_{j+2}\alpha_j) \\ &\quad \times c_3(i+2j+2) + \beta_{j+3}\alpha_j c_3(i+2j+4).\end{aligned} \quad (7.11)$$

Similarly, if we multiply (7.9) by α_{j+2} and (7.10) by α_j, and subtract to eliminate the $c_j(i+j+2)$ term, we have

$$\begin{aligned}(\alpha_j\alpha_{j+1} - \alpha_{j-1}\alpha_{j+2})c_2(i+j+1) &= \beta_{j-1}\alpha_{j+2}c_3(i+2j) - (\beta_j\alpha_{j+2} + \beta_{j+1}\alpha_j) \\ &\quad \times c_3(i+2j+2) + \beta_{j+2}\alpha_j c_3(i+2j+4).\end{aligned}$$

A brief calculation reveals that

$$\alpha_{j+1}\alpha_{j+2} - \alpha_{j+3}\alpha_j = r^2(\alpha_j\alpha_{j+1} - \alpha_{j+2}\alpha_{j-1}),$$

and so we may also eliminate $c_2(i+j+1)$ and conclude that

$$\beta'_j\alpha_{j+2}c_3(i+2j) - (\beta'_{j+1}\alpha_{j+2} + \beta'_{j+2}\alpha_j)c_3(i+2j+2) + \beta'_{j+3}\alpha_j c_3(i+2j+4) = 0 \quad (7.12)$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

for all $-80 \leq i, j \leq 80$, where

$$\beta'_j := \beta_j - r^2\beta_{j-1} = (1 - r^{-2})(r^{4j}t^4 - r^2).$$

We continue the elimination process. Applying (7.12) with (i, j) replaced by $(i + 2, j - 1)$, we conclude that

$$\beta'_{j-1}\alpha_{j+1}c_3(i + 2j) - (\beta'_j\alpha_{j+1} + \beta'_{j+1}\alpha_{j-1})c_3(i + 2j + 2) + \beta'_{j+2}\alpha_{j-1}c_3(i + 2j + 4) = 0$$

for all $-70 \leq i, j \leq 70$ (say). Multiplying this equation by $\beta'_{j+3}\alpha_j$ and (7.12) by $\beta'_{j+2}\alpha_{j-1}$, and subtracting, we conclude that

$$\begin{aligned}(\beta'_{j-1}\beta'_{j+3}\alpha_j\alpha_{j+1} - \beta'_j\beta'_{j+2}\alpha_{j-1}\alpha_{j+2})c_3(i + 2j)\\ = (\beta'_j\beta'_{j+3}\alpha_j\alpha_{j+1} + \beta'_{j+1}\beta'_{j+3}\alpha_{j-1}\alpha_j - \beta'_{j+1}\beta'_{j+2}\alpha_{j-1}\alpha_{j+2}\\ - (\beta'_{j+2})^2\alpha_{j-1}\alpha_j)c_3(i + 2j + 2)\end{aligned}$$

for all $-70 \leq i, j \leq 70$.

We apply this with (i, j) replaced by $(i - 2, 1)$ and $(i - 4, 2)$ to conclude that

$$\begin{aligned}(\beta'_0\beta'_4\alpha_1\alpha_2 - \beta'_1\beta'_3\alpha_0\alpha_3)c_3(i)\\ = (\beta'_1\beta'_4\alpha_1\alpha_2 + \beta'_2\beta'_4\alpha_0\alpha_1 - \beta'_2\beta'_3\alpha_0\alpha_3 - (\beta'_3)^2\alpha_0\alpha_1)c_3(i + 2)\end{aligned}$$

and

$$\begin{aligned}(\beta'_1\beta'_5\alpha_2\alpha_3 - \beta'_2\beta'_4\alpha_1\alpha_4)c_3(i)\\ = (\beta'_2\beta'_5\alpha_2\alpha_3 + \beta'_3\beta'_5\alpha_1\alpha_2 - \beta'_3\beta'_4\alpha_1\alpha_3 - (\beta'_4)^2\alpha_1\alpha_2)c_3(i + 2)\end{aligned}$$

for all $-60 \leq i \leq 60$ (say). Eliminating $c_3(i + 2)$, we conclude that either $c_3(i)$ vanishes for all $-60 \leq i \leq 60$, or else we have the constraint

$$\begin{aligned}(\beta'_0\beta'_4\alpha_1\alpha_2 - \beta'_1\beta'_3\alpha_0\alpha_3)(\beta'_2\beta'_5\alpha_2\alpha_3 + \beta'_3\beta'_5\alpha_1\alpha_2 - \beta'_3\beta'_4\alpha_1\alpha_3 - (\beta'_4)^2\alpha_1\alpha_2)\\ = (\beta'_1\beta'_5\alpha_2\alpha_3 - \beta'_2\beta'_4\alpha_1\alpha_4)(\beta'_1\beta'_4\alpha_1\alpha_2 + \beta'_2\beta'_4\alpha_0\alpha_1\\ - \beta'_2\beta'_3\alpha_0\alpha_3 - (\beta'_3)^2\alpha_0\alpha_1).\end{aligned}$$

After eliminating some factors of $(1 - r^{-2})$, this is a polynomial constraint between r and t of bounded degree. One can easily verify that the constraint is not a tautology (for instance, setting $r = 2$ and $t = 2$, the left-hand side is approximately -1.96×10^{24} and the right-hand side is approximately 3.61×10^{32}). Thus, by the Schwarz–Zippel lemma, there are only $O(|F|)$ possible pairs (r, t), and thus $O(|F|^2)$ triples (r, s, t), that obey this constraint. Outside of those exceptional triples, we thus have $c_3(i)$ vanishing for all $-60 \leq i \leq 60$. Applying (7.11), we conclude that $c_2(0)$ vanishes as well, unless $\alpha_1\alpha_2 - \alpha_3\alpha_0$ vanishes. The latter possibility is also a bounded degree nontautological constraint on r, t, and so it also only occurs for $O(|F|^2)$ triples (r, s, t).

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

Thus we see that $c_3(0)$ and $c_2(0)$ both vanish outside of these exceptional triples. But this contradicts the assumption that η_{i_0} never vanishes (recall that i_0 is either 2 or 3). We have thus demonstrated that there are at most $O(|F|^2)$ triples (r, s, t) for which (7.6) holds for all $-100 \leq i, j \leq 100$. But we also know that there are $\gg |F|^3$ such triples, leading to a contradiction for $|F|$ sufficiently large, as required.

7.2. The contribution of the nonzero frequencies. Finally, we prove (7.3). This will be done by a variant of the Cauchy–Schwarz arguments used to establish Theorem 1.4. Observe that one multiplies $x \in G$ on the left by some element $\psi(k)$ of U; then $f_{i,x}$ and $\Delta_h f_{i,x}$ become translated by k, and the quantity $|\widehat{\Delta_h f_{i,x}}(\xi)|$ is unchanged. Thus, for any $i = 1, 2, 3$, $x \in G$, $h \in F$, and $\xi \in F^\times$, we may write

$$|\widehat{\Delta_h f_{i,x}}(\xi)| = H_{i,h,\pi(x)}(\xi) \tag{7.13}$$

for some function $H_{i,h,\pi(x)}: F^\times \to \mathbf{R}^+$ depending on h and $\pi(x)$. We may thus rewrite (7.3) as

$$\mathbf{E}_{s\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} H_{1,h,s}(\xi)\mathbf{E}_{t\in F^\times}H_{2,(1+t^4)h,st}(-(1+t^{-4})\xi)H_{3,(1+t^4+t^8)h,st^2}(t^{-4}\xi) = o(1).$$

From Plancherel's theorem we have

$$\sum_{\xi\in F^\times} H_{1,h,s}(\xi)^2 \leq 1$$

for all $s \in F^\times$ and $h \in F$, so by the Cauchy–Schwarz inequality it suffices to show that

$$\mathbf{E}_{s\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} |\mathbf{E}_{t\in F^\times}H_{2,(1+t^4)h,st}(-(1+t^{-4})\xi)H_{3,(1+t^4+t^8)h,st^2}(t^{-4}\xi)|^2 = o(1),$$

which we expand as

$$\mathbf{E}_{s,t,u\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} H_{2,(1+t^4)h,st}(-(1+t^{-4})\xi)^2 H_{3,(1+t^4+t^8)h,st^2}(t^{-4}\xi)^2 \times H^2_{2,(1+u^4)h,su}(-(1+u^{-4})\xi)H^2_{3,(1+u^4+u^8)h,su^2}(u^{-4}\xi) = o(1).$$

By another Cauchy–Schwarz inequality and symmetry, it thus suffices to show that

$$\mathbf{E}_{s,t,u\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} H^4_{2,(1+t^4)h,st}(-(1+t^{-4})\xi)H^4_{3,(1+u^4+u^8)h,su^2}(u^{-4}\xi) = o(1).$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

There are at most four values of t for which $t^4 = -1$, and each of these values of t contributes $O(|F|^{-1})$ to the above sum (using Plancherel's theorem $\sum_\xi H_{i,h,s}(\xi) \leq 1$ and the trivial bound $H_{i,h,s}(\xi) \leq 1$), and may be discarded. Dilating h, s, ξ by $(1 + t^4)^{-1}$, t^{-1}, $-(1 + t^{-4})$, respectively, we rewrite the remaining component of the above estimate as

$$\mathbf{E}_{s,t,u\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} 1_{t^4\neq -1}H^4_{2,h,s}(\xi)H^4_{3,(1+u^4+u^8)(1+t^4)^{-1}h,st^{-1}u^2}(-(1+t^{-4})^{-1}u^{-4}\xi) = o(1).$$

Making the change of variables $(s, u, v) := (s, u, st^{-1}u^2)$, so that $t = su^2v^{-1}$, this becomes

$$\mathbf{E}_{s,u,v\in F^\times}\mathbf{E}_{h\in F}\sum_{\xi\in F^\times} 1_{s^4u^8v^{-4}\neq -1}H^4_{2,h,s}(\xi)H^4_{3,(1+u^4+u^8)(1+s^4u^8v^{-4})^{-1}h,v} \times (-(1+s^{-4}u^{-8}v^4)^{-1}u^{-4}\xi) = o(1).$$

From Plancherel's theorem and the trivial bound $H_{2,h,s}(\xi) \leq 1$, we have

$$\sum_{\xi\in F^\times} H^4_{2,h,s}(\xi) \leq 1$$

for each $h \in F$ and $s \in F^\times$. It will thus suffice to establish the bound

$$\mathbf{E}_{u\in F^\times} 1_{s^4u^8v^{-4}\neq -1}H^4_{3,(1+u^4+u^8)(1+s^4u^8v^{-4})^{-1}h,v}(-(1+s^{-4}u^{-8}v^4)^{-1}u^{-4}\xi) = o(1)$$

for all $\xi \in F^\times$, and all but at most $o(|F|^3)$ choices of $(s, v, h) \in F^\times \times F^\times \times F$.

Fix s, v, h. Our task is to show that, for all but $o(|F|^3)$ choices of (s, v, h), one has

$$\mathbf{E}_{u\in F} 1_A(u)H^4_{3,\phi(u),v}(\eta(u))|^4 = o(1), \tag{7.14}$$

where $A := \{u \in F^\times : s^4u^8v^{-4} \neq -1\}$,

$$\phi(u) := (1 + u^4 + u^8)(1 + s^4u^8v^{-4})^{-1}h,$$

and

$$\eta(u) := -(1 + s^{-4}u^{-8}v^4)^{-1}u^{-4}\xi.$$

We may assume that h is nonzero, as this only excludes $O(|F|^2) = o(|F|^3)$ values of (s, v, h).

If we write $f := f_{3,g}$ for some $g \in \pi^{-1}(v)$ and expand the definition (7.13) of $H_{3,h,s}$, we may rewrite (7.14) as

$$\mathbf{E}_{u\in F} 1_A(u)|\widehat{\Delta_{\phi(u)}f}(\eta(u))|^4 = o(1). \tag{7.15}$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

The next step is to apply the Cauchy–Schwarz inequality again, in the spirit of the work of Gowers [9]. First, to show (7.15), it will suffice to show (using the trivial bound $|\widehat{\Delta_h f}(\eta)| \leq 1$) that

$$\mathbf{E}_{u\in F} 1_A(u) |\widehat{\Delta_{\phi(u)} f}(\eta(u))| = o(1),$$

or equivalently that

$$\mathbf{E}_{u\in F} b(u) \widehat{\Delta_{\phi(u)} f}(\eta(u)) = o(1)$$

for any function $b: F \to \mathbf{R}$ supported on A with $|b(u)| \leq 1$ for all u. We can expand the left-hand side as

$$\mathbf{E}_{x,u\in F} b(u) f(x) f(x+\phi(u)) e(-\eta(u)\cdot x),$$

and rearrange this as

$$\mathbf{E}_{x,y\in F} f(x) f(y) K(x,y),$$

where

$$K(x,y) := \sum_{u\in F:\phi(u)=y-x} b(u) e(-\eta(u)\cdot x).$$

Applying the Cauchy–Schwarz inequality twice, and using the boundedness of f, we have

$$|\mathbf{E}_{x,y\in F} f(x) f(y) K(x,y)|^4 \leq \mathbf{E}_{x,y,x',y'\in F} K(x,y)\overline{K(x,y')K(x',y)}K(x',y'),$$

so it will suffice to show that

$$\mathbf{E}_{x,y,x',y'\in F} \overline{K(x,y')K(x',y)}K(x',y') = o(1).$$

The left-hand side may be expanded as

$$\begin{aligned} |F|^{-4} &\sum_{u_1,u_2,u_3,u_4\in A} b(u_1)b(u_2)b(u_3)b(u_4) \\ &\times \sum_{x,y,x',y'\in F:\phi(u_1)=x-y,\phi(u_2)=x-y',\phi(u_3)=x'-y,\phi(u_4)=x'-y'} \\ &\times e(-(\eta(u_1)-\eta(u_2)-\eta(u_3)+\eta(u_4))\cdot x) e((\eta(u_3)-\eta(u_4))\cdot(x'-x)). \end{aligned}$$

The quantity $x'-x$ in the summand is equal to $\phi(u_3)-\phi(u_1)$, and so this phase is constant over the inner summation. By Fourier analysis, we see that the inner summation is thus $O(|F|)$ when $\eta(u_1)+\eta(u_4) = \eta(u_2)+\eta(u_3)$ and $\phi(u_1)+\phi(u_4) = \phi(u_2)+\phi(u_3)$, and zero otherwise. It thus suffices to show that

$$\begin{aligned} |\{(u_1,u_2,u_3,u_4)\in A^4 : \eta(u_1)+\eta(u_4) = \eta(u_2)+\eta(u_3); \phi(u_1)+\phi(u_4) \\ = \phi(u_2)+\phi(u_3)\}| = o(|F|^3). \end{aligned}$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

Canceling out the nonzero h and ξ factors, and replacing each of the u_i by their fourth powers (at the cost of paying $O(1)$ in the cardinality bound), this becomes

$$|\{(u_1, u_2, u_3, u_4) \in A^4 : \Phi(u_1) + \Phi(u_4) = \Phi(u_2) + \Phi(u_3)\}| = o(|F|^3),$$

where $\Phi : F \to F^2$ is the rational function

$$\Phi(u) := ((1 + u + u^2)(1 + ku^2)^{-1}, (1 + k^{-1}u^{-2})^{-1}u^{-1})$$

and $k := s^4v^{-4}$. We can simplify $(1 + k^{-1}u^{-2})^{-1}u^{-1}$ as $ku(1 + ku^2)^{-1}$ and $(1 + u + u^2)(1 + ku^2)^{-1}$ as $k^{-1} + (1 - k^{-1} + u)(1 + ku^2)^{-1}$, so, after excluding the $O(|F|^2) = o(|F|^3)$ triplets (s, v, h) for which $k = 1$, we may replace Φ by

$$\tilde{\Phi}(u) := ((1 + ku^2)^{-1}, u(1 + ku^2)^{-1}).$$

This function takes values in the conic section

$$C := \{(x, y) \in F : x^2 + ky^2 = x\}$$

with each point in C arising from at most two values of u, and so it suffices to show that

$$|\{(p_1, p_2, p_3, p_4) \in C^4 : p_1 + p_4 = p_2 + p_3\}| = o(|F|^3).$$

But from Bezout's theorem we see that each point in F^2 can be expressed in at most two ways as the sum of two elements in C, and so the left-hand side is $O(|F|^2)$, and the claim follows.

REMARK 7.3. The above argument in fact allows us to replace $o(1)$ by $O(|F|^{-c})$ for some absolute constant $c > 0$, for the contribution of the nonzero frequencies ξ. Unfortunately, due to the reliance on the multidimensional Szemerédi theorem, we are unable to obtain a similarly strong bound for the contribution of the zero frequencies.

Acknowledgements

The author was partially supported by a Simons Investigator award from the Simons Foundation and by NSF grant DMS-0649473. He also thanks Vitaly Bergelson for many stimulating discussions regarding these topics.

Appendix A. Some algebraic geometry

Throughout this appendix, k is an algebraically closed field, and F is a finite subfield of k. The purpose of this appendix is to review some basic algebraic geometry regarding varieties and regular maps over k.

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

We begin with the definition of a variety. For the purposes of this paper, we may restrict attention to affine varieties for simplicity, but most of the results here can be extended to other types of variety (projective, quasiprojective, and so on).

DEFINITION A.1 (Varieties). An *(affine) variety* defined over k is a subset $V \subseteq k^n$ of the form

$$V = \{x \in k^n : P_1(x) = \cdots = P_m(x) = 0\},$$

where n, m are natural numbers, and $P_1, \ldots, P_m : k^n \to k$ are polynomials. We say that the variety has *complexity at most* M if n, m are at most M, and all the degrees of $P_1, \ldots, P_m$ are at most M. If, furthermore, the polynomials $P_1, \ldots, P_m$ have coefficients defined over F, we say that V is *defined over* F (with complexity at most M). A variety is *(geometrically) irreducible* if it cannot be expressed as the union of two strictly smaller subvarieties.

The *Zariski closure* of a subset E of k^n is defined to be the intersection of all the varieties in k^n that contain E.

The *dimension* of a nonempty variety $V \subset k^n$ is the largest natural number d for which one has a chain

$$\emptyset \subsetneq V_0 \subsetneq \cdots \subsetneq V_d \subset V$$

of irreducible varieties $V_0, \ldots, V_d$. We adopt the convention that the empty set has dimension $-\infty$.

We have the following basic upper bound for the number of F-points on a variety.

PROPOSITION A.2 (Schwarz–Zippel bound). *Let $V \subset k^m$ be an affine variety defined over k of complexity at most M and dimension d. Then*

$$|V \cap F^m| \ll_{m,M} |F|^d.$$

Proof. See for instance [17, Lemma 1]. One can make the implied constant depend linearly on the degree of V, but we will not need this refinement here. □

In the case that V is irreducible and defined over F, we have the following well-known refinement of Proposition A.2.

PROPOSITION A.3 (Lang–Weil bound). *Let $V \subset k^m$ be a geometrically irreducible affine variety defined over F of complexity at most M and dimension d. Then*

$$|V \cap F^m| = (1 + O_{m,M}(|F|^{-1/2}))|F|^d.$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

In particular, if $|F|$ is sufficiently large depending on m, M, one has

$$|F|^d \ll |V \cap F^m| \ll |F|^d.$$

Proof. See [17, Theorem 1]. Again, more precise versions of the error term are available, but we will not need them here. □

Now we recall the notions of regular and dominant maps between varieties. Our definition will be somewhat complicated due to the need to assign quantitative complexities to such maps.

DEFINITION A.4 (Regular map). Let $V \subset k^n$ and $W \subset k^m$ be affine varieties, and let $M \geq 1$. A map $f: V \to W$ is said to be *regular* with complexity at most M if V, W are individually of complexity at most M, and if one can cover V by some varieties $V_1, \ldots, V_r$ of complexity at most M for some $r \leq M$ such that, for each $1 \leq j \leq r$, the map $f|_{V_j}$ has the form $(P_{j,1}/Q_{j,1}, \ldots, P_{j,m}/Q_{j,m})$, where the $P_{j,l}, Q_{j,l}$ are homogeneous polynomial maps from k^{n+1} to k with $\deg(P_{j,l}) = \deg(Q_{j,l}) \leq M$, and the $Q_{j,l}$ are nonvanishing on V_j.

A regular map $\phi: V \to W$ is *dominant* if V is irreducible and $\phi(V)$ is Zariski dense in W.

The following proposition asserts (in a certain technical quantitative sense) that regular maps are always 'essentially dominant' after a reduction in the range, and that the fibres of such maps usually have the expected dimension.

PROPOSITION A.5 (Quantitative dominance). *Let $V \subset k^m$, $W \subset k^n$ be algebraic varieties defined over k of complexity at most M, with V irreducible, and let $\phi: V \to W$ be a regular map of complexity at most M. Then there exists a subset V' of V and an irreducible subvariety W' of W of complexity $O_M(1)$, with the following properties.*

(i) *(Zariski density) $V \backslash V'$ can be covered by the union of $O_M(1)$ varieties of complexity $O_M(1)$ and dimension strictly less than* $\dim(V)$.

(ii) *(Controlled image) W' is equal to the Zariski closure of $\phi(V)$; in particular, $\phi: V \to W'$ is a dominant map.*

(iii) *(Controlled fibres) For each $w \in W'$, the set $\{v \in V' : \phi(v) = w\}$ can be covered by the union of $O_M(1)$ varieties of complexity $O_M(1)$ and dimension at most* $\dim(V) - \dim(W')$.

Proof. This follows from [7, Lemma 3.7]. □

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

Appendix B. A quantitative multidimensional Szemerédi theorem

The purpose of this section is to establish the following multidimensional Szemerédi theorem.

THEOREM B.1 (Multidimensional Szemerédi theorem). *Let $G = (G, +)$ be an additive group, let $k, m \geq 1$ be integers, and let $A \subset G^m$ be a set with $|A| \geq \delta |G|^m$. Then there are $\gg_{k,m,\delta} |G|^{m+1}$ tuples $(a_1, \ldots, a_m, r) \in G^{m+1}$ with the property that*

$$(a_1 + i_1 r, \ldots, a_m + i_m r) \in A$$

for all integers $i_1, \ldots, i_m \in \{-k, \ldots, k\}$.

This is a variant of the multidimensional Szemerédi theorem of Furstenberg and Katznelson [8]. There are now many techniques to establish such results; we will derive Theorem B.1 from the hypergraph removal lemma established in [11, 22, 23, 27].

We first observe that Theorem B.1 may be deduced via a lifting trick from the following apparently weaker version.

THEOREM B.2 (Multidimensional Szemerédi theorem, again). *Let $G = (G, +)$ be an additive group, let $m \geq 1$ be integers, and let $A \subset G^m$ be a set with $|A| \geq \delta |G|^m$. Then there are $\gg_{m,\delta} |G|^{m+1}$ tuples $(a, r) \in G^m \times G$ with the property that*

$$a + re_1, \ldots, a + re_m \in A$$

where we adopt the notation that $g(n_1, \ldots, n_m) := (n_1 g, \ldots, n_m g)$ whenever $g \in G$ and $n_1, \ldots, n_m$ are integers, and $e_1, \ldots, e_m$ is the standard basis of $\mathbf{Z}^m$.

Indeed, to deduce Theorem B.1 from Theorem B.2, let $K := (2k+1)^m$, and let $v_1, \ldots, v_K$ be an enumeration of the Km-tuples in $\{-k, \ldots, k\}^m$. If $A \subset G^m$, we let $\tilde{A} \subset G^{m+K}$ be the set

$$\tilde{A} := \{(a, b_1, \ldots, b_K) \in G^m \times G^K : a + b_1 v_1 + \cdots + b_K v_K \in A\}.$$

If $|A| \geq \delta |G|^m$, then it is clear (by freezing $b_1, \ldots, b_K$) that $|\tilde{A}| \geq \delta |G|^{m+K}$. Applying Theorem B.2, we see that there are $\gg_{k,m,\delta}$ tuples $(a, b_1, \ldots, b_K, r) \in G^{m+K+1}$ such that

$$(a, b_1, \ldots, b_{i-1}, b_i + r, b_{i+1}, \ldots, b_K) \in \tilde{A}$$

for all $1 \leq i \leq K$, which by the definition of $\tilde{A}$ implies that

$$a' + r v_i \in A \tag{B 1}$$

for all $i = 1, \ldots, K$, where $a' := a + b_1 v_1 + \cdots + b_K v_K$. Since each $a' \in G^m$ arises from at most $|G|^K$ tuples $(a, b_1, \ldots, b_K)$, we conclude that there are $\gg_{k,m,\delta}$ tuples $(a', r) \in G^{m+1}$ such that (B 1) holds for all $i = 1, \ldots, K$, and the claim follows.

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

We now establish Theorem B.2. Let G, A, m be as in that theorem. For each $i = 1, \dots, m$, we introduce a set $E_i \subset G^{m+1}$, defined as the set of all tuples $(a_1, \dots, a_m, s) \in G^{m+1}$ with the property that

$$(a_1, \dots, a_{i-1}, s - a_1 - \cdots - a_{i-1} - a_{i+1} - \cdots - a_m, a_{i+1}, \dots, a_m) \in A.$$

Observe that, if $(a_1, \dots, a_m, s)$ lies in the intersection $\bigcap_{i=1}^m E_i$ of all the E_i, then, by setting $r := s - a_1 - \cdots - a_m$, we have $(a_1, \dots, a_m) + re_i \in A$ for all $i = 1, \dots, m$. Thus it will suffice to show that

$$\left|\bigcap_{i=1}^m E_i\right| \gg_{m,\delta} |G|^{m+1}.$$

Let $\varepsilon > 0$ be a sufficiently small quantity depending on m, δ to be chosen later. Suppose for sake of contradiction that

$$\left|\bigcap_{i=1}^m E_i\right| < \varepsilon |G|^{m+1}.$$

Observe that each E_i is *i-invariant* in the sense that the assertion that a given tuple $(a_1, \dots, a_m, s) \in G^{m+1}$ lies in E_i does not depend on the ith coordinate a_i. Because of this, we may apply the hypergraph removal lemma (see for example [27, Theorem 1.13]) and conclude (if ε is small enough depending on m, δ) that there exist i-invariant perturbations E_i' of E_i with

$$|E_i' \Delta E_i| < \frac{\delta}{m} |G|^{m+1} \tag{B 2}$$

such that

$$\bigcap_{i=1}^m E_i' = \emptyset. \tag{B 3}$$

We now intersect E_i, E_i' with the hyperplane

$$\Sigma := \{(a_1, \dots, a_m, a_1 + \cdots + a_m) : a_1, \dots, a_m \in G\}.$$

As this hyperplane sits transversely with respect to the i-invariant set $E_i' \Delta E_i$, we conclude from (B 2) that

$$|(E_i' \Delta E_i) \cap \Sigma| < \frac{\delta}{m} |G|^m,$$

and hence, from the union bound and (B 3),

$$\left|\bigcap_{i=1}^m E_i \cap \Sigma\right| < \delta |G|^m.$$

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

On the other hand, since $(a_1, \dots, a_m, a_1 + \cdots + a_m) \in \bigcap_{i=1}^m E_i \cap \Sigma$ whenever $(a_1, \dots, a_m) \in A$, we have

$$\left| \bigcap_{i=1}^m E_i \cap \Sigma \right| \geq |A| \geq \delta |G|^m,$$

giving the desired contradiction. This completes the proof of Theorem B.2, and hence Theorem B.1.

References

[1] L. Babai, N. Nikolov and L. Pyber, 'Product growth and mixing in finite groups', In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (ACM, New York, 2008), 248–257.

[2] M. Bateman and N. H. Katz, 'New bounds on cap sets', *J. Amer. Math. Soc.* **25** (2012), 585–613.

[3] F. A. Behrend, 'On sets of integers which contain no three terms in arithmetic progression', *Proc. Natl. Acad. Sci.* **32** (1946), 331–332.

[4] V. Bergelson and T. Tao, 'Multiple recurrence in quasirandom groups', to appear. arXiv:1211.6372.

[5] J. Bourgain, 'On triples in arithmetic progression', *Geom. Funct. Anal.* **9** (1999), 968–984.

[6] J. Bourgain and A. Gamburd, 'Uniform expansion bounds for Cayley graphs of $\mathrm{SL}_2(F_p)$', *Ann. of Math.* (2) **167**(2) (2008), 625–642.

[7] E. Breuillard, B. Green and T. Tao, 'Approximate subgroups of linear groups', *Geom. Funct. Anal.* **21** (2011), 774–819.

[8] H. Furstenberg and Y. Katznelson, 'An ergodic Szemerédi theorem for commuting transformations', *J. Anal. Math.* **34** (1978), 275–291 (1979).

[9] W. T. Gowers, 'A new proof of Szemerédi's theorem for progressions of length four', *Geom. Funct. Anal.* **8**(3) (1998), 529–551.

[10] W. T. Gowers, 'A new proof of Szemerédi's theorem', *Geom. Funct. Anal.* **11**(3) (2001), 465–588.

[11] W. T. Gowers, 'Hypergraph regularity and the multidimensional Szemerédi theorem', *Ann. of Math.* (2) **166**(3) (2007), 897–946.

[12] W. T. Gowers, 'Quasirandom groups', *Combin. Probab. Comput.* **17**(3) (2008), 363–387.

[13] B. Green and T. Tao, 'New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$', In *Analytic Number Theory* (Cambridge University Press, Cambridge, 2009), 180–204.

[14] H. A. Helfgott, 'Growth and generation in $\mathrm{SL}_2(\mathbf{Z}/p\mathbf{Z})$', *Ann. of Math.* (2) **167**(2) (2008), 601–623.

[15] J. Humphreys, *Linear Algebraic Groups*, Graduate Texts in Mathematics, 21 (Springer, New York-Heidelberg, 1975).

[16] V. Landazuri and G. Seitz, 'On the minimal degrees of projective representations of the finite Chevalley groups', *J. Algebra* **32** (1974), 418–443.

[17] S. Lang and A. Weil, 'Number of points of varieties in finite fields', *Amer. J. Math.* **76** (1954), 819–827.

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press

[18] A. Lubotzky, 'Expander graphs in pure and applied mathematics', *Bull. Amer. Math. Soc. (N.S.)* **49**(1) (2012), 113–162.

[19] D. H. J. Polymath, 'A new proof of the density Hales–Jewett theorem', *Ann. of Math.* (2) **175**(3) (2012), 1283–1327.

[20] L. Pyber and E. Szabó, 'Growth in finite simple groups of Lie type', Preprint (2010) arXiv:1001.4556.

[21] R. A. Rankin, 'Sets of integers containing not more than a given number of terms in arithmetical progression', *Proc. Roy. Soc. Edinburgh Sect.* A **65** (1960/1961), 332–344.

[22] V. Rödl and M. Schacht, 'Regular partitions of hypergraphs: regularity lemmas', *Combin. Probab. Comput.* **16**(6) (2007), 833–885.

[23] V. Rödl and J. Skokan, 'Applications of the regularity lemma for uniform hypergraphs', *Random Structures Algorithms* **28**(2) (2006), 180–194.

[24] T. Sanders, 'On Roth's theorem on progressions', *Ann. of Math.* (2) **174**(1) (2011), 619–636.

[25] J. Schwartz, 'Fast probabilistic algorithms for verification of polynomial identities', *J. ACM* **27** (1980), 701–717.

[26] E. Szemerédi, 'On sets of integers containing no k elements in arithmetic progression', *Acta Arith.* **27** (1975), 299–345.

[27] T. Tao, 'A variant of the hypergraph removal lemma', *J. Combin. Theory Ser.* A **113**(7) (2006), 1257–1280.

[28] T. Tao, 'Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets', Preprint. arXiv:1211.2894.

[29] T. Tao and V. Vu, *Additive Combinatorics* (Cambridge University Press, 2006).

https://doi.org/10.1017/fms.2013.2 Published online by Cambridge University Press