Contributions to Discrete Mathematics

Volume 4, Number 2, Pages 1–36
ISSN 1715-0868

THE DISTRIBUTION OF POLYNOMIALS OVER FINITE FIELDS, WITH APPLICATIONS TO THE GOWERS NORMS

BEN GREEN AND TERENCE TAO

Abstract. In this paper we investigate the uniform distribution properties of polynomials in many variables and bounded degree over a fixed finite field $\mathbb{F}$ of prime order. Our main result is that a polynomial $P : \mathbb{F}^n \to \mathbb{F}$ is poorly-distributed only if P is determined by the values of a few polynomials of lower degree, in which case we say that P has *small rank*.

We give several applications of this result, paying particular attention to consequences for the theory of the so-called *Gowers norms*. We establish an inverse result for the Gowers U^{d+1}-norm of functions of the form $f(x) = e_{\mathbb{F}}(P(x))$, where $P : \mathbb{F}^n \to \mathbb{F}$ is a polynomial of degree less than $|\mathbb{F}|$, showing that this norm can only be large if f correlates with $e_{\mathbb{F}}(Q(x))$ for some polynomial $Q : \mathbb{F}^n \to \mathbb{F}$ of degree at most d.

The requirement $\deg(P) < |\mathbb{F}|$ cannot be dropped entirely. Indeed, we show the above claim fails in characteristic 2 when $d = 3$ and $\deg(P) = 4$, showing that the quartic symmetric polynomial S_4 in $\mathbb{F}_2^n$ has large Gowers U^4-norm but does not correlate strongly with any cubic polynomial. This shows that the theory of Gowers norms in low characteristic is not as simple as previously supposed. This counterexample has also been discovered independently by Lovett, Meshulam, and Samorodnitsky [15].

We conclude with sundry other applications of our main result, including a recurrence result and a certain type of nullstellensatz.

1. Introduction

Let $\mathbb{F}$ be a finite field of prime order. Throughout this paper, $\mathbb{F}$ will be considered fixed (e.g. $\mathbb{F} = \mathbb{F}_2$ or $\mathbb{F} = \mathbb{F}_3$) and we shall be working inside the n-dimensional vector spaces $\mathbb{F}^n$ over $\mathbb{F}$ for various natural numbers n. More generally, any linear algebra term (e.g. span, independence, basis, subspace, linear transformation, etc.) will be understood to be over the field $\mathbb{F}$.

Received by the editors November 20, 2007, and in revised form February 13, 2009.
2000 *Mathematics Subject Classification.* (Missing.)
Key words and phrases. (Missing.)
The first author is a Clay Research Fellow and gratefully acknowledges the support of the Clay Institute. The second author is supported by a grant from the MacArthur Foundation, and by NSF grant CCF-0649473.

If $f : \mathbb{F}^n \to \mathbb{C}$ is a function, and $h \in \mathbb{F}^n$ is a shift, we define the (multiplicative) derivative $\Delta_h f : \mathbb{F}^n \to \mathbb{C}$ of f by the formula

$$\Delta_h f(x) := f(x+h)\overline{f(x)}.$$

An important special case arises when f takes the form $f = e_{\mathbb{F}}(P)$, where $P : \mathbb{F}^n \to \mathbb{F}$ is a function, and $e_{\mathbb{F}} : \mathbb{F} \to \mathbb{C}$ is the standard character $e_{\mathbb{F}}(j) := e^{2\pi ij/|\mathbb{F}|}$ for $j = 0, \dots, |\mathbb{F}| - 1$. In that case we see that $\Delta_h f = e_{\mathbb{F}}(D_h P)$, where $D_h P : \mathbb{F}^n \to \mathbb{F}$ is the (additive) derivative of P, defined as

$$D_h P(x) := P(x+h) - P(x).$$

Given an integer $d \geqslant 0$, we say that a function $P : \mathbb{F}^n \to \mathbb{F}$ is a *polynomial of degree at most d* if we have $D_{h_1} \cdots D_{h_{d+1}} P = 0$ for all $h_1, \dots, h_{d+1} \in \mathbb{F}^n$, and write $\mathcal{P}_d(\mathbb{F}^n)$ for the space of all polynomials on $\mathbb{F}^n$ of degree at most d. Thus for instance $\mathcal{P}_0(\mathbb{F}^n)$ is the space of constants, $\mathcal{P}_1(\mathbb{F}^n)$ is the space of linear polynomials on $\mathbb{F}^n$, $\mathcal{P}_2(\mathbb{F}^n)$ is the space of quadratic polynomials, and so forth. It is easy to see that $\mathcal{P}_d(\mathbb{F}^n)$ is a vector space and that, with an obvious notation, the monomials $x_1^{i_1} \cdots x_n^{i_n}$ for $0 \leqslant i_1, \dots, i_n < |\mathbb{F}|$ and $i_1 + \cdots + i_n \leqslant d$ form a basis. (The restriction $i_1, \dots, i_n < |\mathbb{F}|$ arises of course from the fact that $x^{|\mathbb{F}|} = x$ for all $x \in \mathbb{F}$.) We shall say that a function $f : \mathbb{F}^n \to \mathbb{C}$ is a *polynomial phase of degree at most d* if it takes the form $f = e_{\mathbb{F}}(P)$ for some $P \in \mathcal{P}_d(\mathbb{F}^n)$, or equivalently if all $(d+1)^{\text{st}}$ multiplicative derivatives $\Delta_{h_1} \cdots \Delta_{h_{d+1}} f$ are identically 1.

It is of interest to test for the property that a function $P : \mathbb{F}^n \to \mathbb{F}$ is "close to" a polynomial of degree at most d, or to test for the closely related property that a function $f : \mathbb{F}^n \to \mathbb{C}$ "correlates" with a polynomial phase of degree at most d. One proposal to perform such a test goes by the name of the *Inverse Conjecture for the Gowers norms* (see e.g. [6, 12, 18]), which roughly speaking asserts that a function f correlates with a polynomial phase of degree at most d if and only if the $(d+1)^{\text{st}}$ multiplicative derivatives of f are biased. To describe this conjecture more precisely, we need some further notation.

Definition 1.1 (Gowers uniformity norm [8], [9])**.** *Let $f : \mathbb{F}^n \to \mathbb{C}$ be a function, and let $d \geqslant 0$ be an integer. We then define the* Gowers norm $\|f\|_{U^{d+1}}$ *of f to be the quantity*[1]

$$\|f\|_{U^{d+1}} := \left|\mathbb{E}_{h_1,\dots,h_d,x \in \mathbb{F}^n} \Delta_{h_1} \cdots \Delta_{h_{d+1}} f(x)\right|^{1/2^{d+1}},$$

thus $\|f\|_{U^{d+1}}$ measures the average bias in $(d+1)^{\text{st}}$ multiplicative derivatives of f. We also define the weak Gowers norm $\|f\|_{u^{d+1}}$ *of f to be the quantity*

$$\|f\|_{u^{d+1}} := \sup_{Q \in \mathcal{P}_d(\mathbb{F}^n)} |\mathbb{E}_{x \in \mathbb{F}^n} f(x) e_{\mathbb{F}}(-Q(x))|,$$

thus $\|f\|_{u^{d+1}}$ measures the extent to which f can correlate with a polynomial phase of degree at most d.

[1]Here, as in all our papers, the expectation notation $\mathbb{E}_{x \in S}$ refers to the average $\frac{1}{|S|}\sum_{x \in S}$ over some finite non-empty set S. In this particular example, $S = (\mathbb{F}^n)^{d+1}$.

Remark: It can in fact be shown that the Gowers and weak Gowers norm are in fact norms for $d \geqslant 2$ (and seminorms for $d = 1$), see e.g. [9, 19]. Further discussion of these two norms can be found in [12].

The Gowers norm and weak Gowers norm are closely related; for instance, one easily verifies the invariance

$$\|fg\|_{U^{d+1}} = \|f\|_{U^{d+1}} \text{ and } \|fg\|_{u^{d+1}} = \|f\|_{u^{d+1}} \tag{1.1}$$

for all functions $f : \mathbb{F}^n \to \mathbb{C}$ and all polynomial phases g of degree at most d, and from this and the Cauchy-Schwarz-Gowers inequality (see e.g. [19]) one can also verify the bound

$$\|f\|_{u^{d+1}} \leqslant \|f\|_{U^{d+1}} \tag{1.2}$$

whenever f is bounded in magnitude by 1. In the converse direction the following had been suggested, and is stated formally[2] in [16, 18].

Conjecture 1.3 (Inverse conjecture for the Gowers norm)**.** *Let $d \geqslant 0$, let $\delta \in (0, 1]$, and $\mathbb{F}$ be a fixed finite field. Suppose that $f : \mathbb{F}^n \to \mathbb{C}$ is a function with $|f(x)| \leqslant 1$ for all $x \in \mathbb{F}^n$ and for which $\|f\|_{U^{d+1}} \geqslant \delta$. Then $\|f\|_{u^{d+1}} \gg_{d,\delta,\mathbb{F}} 1$; that is to say, there is some $c = c(d, \delta, \mathbb{F}) > 0$ such that $\|f\|_{u^{d+1}} \geqslant c$.*

This conjecture has been verified in a number of special cases. For instance the case $d = 0$ is trivial, and the case $d = 1$ is easily established by Plancherel's theorem. The case $d = 2$ is established for fields of odd characteristic in [12] and in the case $|\mathbb{F}| = 2$ (which is of particular interest in theoretical computer science) in [16]. The case when δ is sufficiently close to 1 (depending on d and $\mathbb{F}$) is established in [3] (see also the earlier related work of [5] in the case $d = 1$, and [17] in the case when $|\mathbb{F}|$ is assumed large compared to d and δ).

One of our results in this paper establishes a further special case of the conjecture, when the function f is itself a polynomial phase, and the characteristic of $\mathbb{F}$ is not too small.

Theorem 1.4 (Inverse conjecture for polynomial phases)**.** *Suppose that $0 \leqslant d, k < |\mathbb{F}|$, and that $\delta \in (0, 1]$. Let $P : \mathbb{F}^n \to \mathbb{F}$ be a polynomial of degree k, write $f(x) := e_{\mathbb{F}}(P(x))$, and suppose that $\|f\|_{U^{d+1}} \geqslant \delta$. Then we have $\|f\|_{u^{d+1}} \gg_{\mathbb{F},\delta} 1$.*

Note carefully the lower bound on the characteristic $|\mathbb{F}|$ of $\mathbb{F}$. It turns out that some such restriction is necessary, and indeed that Conjecture 1.3 is false without some modification. This is elucidated by the following example, which we shall analyse in §10. For any $d \geqslant 0$ and any vector space $\mathbb{F}^n$, let $S_d \in \mathcal{P}_d(\mathbb{F}^n)$ be the symmetric polynomial of degree d:

$$S_d(x_1, \ldots, x_n) := \sum_{1 \leqslant i_1 < \cdots < i_d \leqslant n} x_{i_1} \cdots x_{i_d}. \tag{1.3}$$

[2]The first-named author would like to make it clear that he also believed the conjecture.

Theorem 1.5 (Counterexample for the U^4-norm in $\mathbb{F}_2$)**.** *Let n be a large integer. Then the function $f : \mathbb{F}_2^n \to \{-1, 1\}$ defined by $f := e_{\mathbb{F}_2}(S_4) = (-1)^{S_4}$ is such that*

$$\|f\|_{U^4}^{16} = \frac{1}{8} + O(2^{-n/2}) \tag{1.4}$$

but such that

$$\|f\|_{u^4} \ll (\log n)^{-c} \tag{1.5}$$

for some absolute constant $c > 0$.

This counterexample was discovered independently by Lovett, Meshulam and Samorodnitsky [15]. They obtain a very much stronger bound for the lack of correlation of f with a cubic phase, namely $\|f\|_{u^4} \ll 2^{-cn}$. We obtain our bound by a very slight modification of Ramsey-theoretic arguments of Alon and Beigel [2]. We will in fact be able to establish similar results with S_4 replaced by S_{2^j} for $j \geqslant 2$; see Theorem 11.3. The aforementioned paper of Lovett, Meshulam and Samorodnitsky goes further in establishing counterexamples to Conjecture 1.3 for all prime fields $\mathbb{F} = \mathbb{F}_p$; specifically, the conjecture fails when $d + 1 = p^2$.

We note that the counterexample presented in Theorem 1.5 is also a counterexample to the specific case of Conjecture 1.3 given as [6, Conjecture 21].

It seems of interest to determine for what other degrees, Gowers norms, and characteristics one has a counterexample of the above type, and to ask what can be salvaged when $\mathbb{F}$ is very small. We will speculate on these questions in §11. We do not regard Theorem 1.5 as an obstacle to the possible truth of the inverse conjecture over $\mathbb{Z}/N\mathbb{Z}$ on which our programme to count solutions to linear equations in primes depends (cf. [11]). Indeed this seems to be a "low characteristic" issue, albeit one of a rather interesting nature.

We turn now to a discussion of the main technical result of the paper, on which the proof of Theorem 1.4 depends. We begin by defining the notion of *rank*.

Definition 1.6 (Rank)**.** *Let $d \geqslant 0$, and let $P : \mathbb{F}^n \to \mathbb{F}$ be a function. We define the* degree d rank $\mathrm{rank}_d(P)$ *of P to be the least integer $k \geqslant 0$ for which there exist polynomials $Q_1, \dots, Q_k \in \mathcal{P}_d(\mathbb{F}^n)$ and a function $B : \mathbb{F}^k \to \mathbb{F}$ such that we have the representation $P = B(Q_1, \dots, Q_k)$. If no such k exists, we declare $\mathrm{rank}_d(P)$ to be infinite (since $\mathbb{F}^n$ is finite-dimensional, this only occurs when $d = 0$ and P is non-constant).*

In the low-degree case, it is well known that the bias $\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))$ of a polynomial phase $e_{\mathbb{F}}(P(x))$ is closely related to the rank of P. For instance, if $P \in \mathcal{P}_1(\mathbb{F}^n)$ is linear, then from simple Fourier analysis we see that $\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))$ has magnitude 1 if $\mathrm{rank}_0(P) = 0$ and magnitude 0 otherwise. For quadratic polynomials, we have the following well-known fact:

Lemma 1.7 (Gauss sum estimate)**.** *If $P \in \mathcal{P}_2(\mathbb{F}^n)$, then*

$$|\mathbb{E}_{x\in\mathbb{F}^n} e_{\mathbb{F}}(P(x))| \ll |\mathbb{F}|^{-c\,\mathrm{rank}_1(P)}$$

where $c > 0$ is an absolute constant.

Proof. If $P \in \mathcal{P}_1(\mathbb{F}^n)$ then the claim can be verified by Fourier analysis, so we can assume that $P \notin \mathcal{P}_1(\mathbb{F}^n)$. We begin with the easy case $|\mathbb{F}| > 2$, and then discuss the changes needed to handle $|\mathbb{F}| = 2$.

Suppose that

$$|\mathbb{E}_{x\in\mathbb{F}^n} e_{\mathbb{F}}(P(x))| \geqslant \delta \tag{1.6}$$

for some $0 < \delta < 1/2$. It will suffice to show that $\mathrm{rank}_1(P) \ll \log_{|\mathbb{F}|} \frac{1}{\delta}$.

Squaring (1.6), we conclude that

$$\delta^2 \leqslant \mathbb{E}_{x,y\in\mathbb{F}^n} e_{\mathbb{F}}(P(x) - P(y)) = \mathbb{E}_{x,h\in\mathbb{F}^n} e_{\mathbb{F}}(D_h P(x)).$$

From Fourier analysis, we see that the average $\mathbb{E}_{x\in\mathbb{F}^n} e_{\mathbb{F}}(D_h P(x))$ vanishes unless $D_h P \in \mathcal{P}_0(\mathbb{F}^n)$, in which case it has magnitude 1. Thus the assumption (1.6) implies that

$$\mathbb{P}_{h\in\mathbb{F}^n}(D_h P \in \mathcal{P}_0(\mathbb{F}^n)) \geqslant \delta^2.$$

Now by breaking up P into monomials, we can express $P(x) = B(x,x)+L(x)$ for some bilinear form $B : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{F}$ and some $L \in \mathcal{P}_1(\mathbb{F}^n)$. Indeed if $P(x) = \sum_{i\leqslant j} a_{ij}x_i x_j + \sum_i b_i x_i + c$ then we may take $B(x,y) = \sum_{i\leqslant j} a_{ij}x_i y_j$ and $L(x) = \sum_i b_i x_i + c$. In the odd characteristic case $|\mathbb{F}| > 2$, we can take B to be symmetric by setting $B(x,y) = \sum_{i\leqslant j} \frac{1}{2}a_{ij}x_j y_i + \sum_{j\leqslant i} \frac{1}{2}a_{ji}x_i y_j$. We conclude that

$$D_h P(x) = 2B(x,h) \pmod{\mathcal{P}_0(\mathbb{F}^n)},$$

and hence that

$$\mathbb{P}_{h\in\mathbb{F}^n}(B(x,h) = 0 \text{ for all } x \in \mathbb{F}^n) \geqslant \delta^2.$$

If $\delta^2 > 1/|\mathbb{F}|$ then, since the set $\{h \in \mathbb{F}^n : B(x,h) = 0 \text{ for all } x \in \mathbb{F}^n\}$ is a vector subspace, the form B vanishes identically. This contradicts the hypothesis $P \notin \mathcal{P}_1(\mathbb{F}^n)$, so we may assume $\delta^2 \leqslant 1/|\mathbb{F}|$. Then the linear transformation associated to B has rank at most $O(\log_{|\mathbb{F}|} 1/\delta)$; since $P(x) = B(x,x) + L(x)$, we conclude $\mathrm{rank}_1(P) \ll \log_{|\mathbb{F}|} 1/\delta$ as desired.

Now we consider the even characteristic case $|\mathbb{F}| = 2$, in which case we cannot take B to be symmetric. Then the above argument gives

$$\mathbb{P}_{h\in\mathbb{F}^n}(\tilde{B}(x,h) = 0 \text{ for all } x \in \mathbb{F}^n) \geqslant \delta^2,$$

where $\tilde{B}(x,h) := B(x,h) + B(h,x)$ is a symmetric bilinear form. Thus $\tilde{B}$ must have rank $O(\log_2 1/\delta)$. By linear algebra we can thus express

$$\tilde{B}(x,h) = \sum_{1\leqslant i,j\leqslant k} c_{i,j} L_i(x) L_j(h),$$

for some $k \ll \log_2 1/\delta$, some linearly independent linear functionals $L_i : \mathbb{F}^n \to \mathbb{F}$, and some coefficients $c_{i,j} \in \mathbb{F}$. Since $\tilde{B}$ is symmetric and the L_i are

independent, we have $c_{i,j} = c_{j,i}$. Since $\tilde{B}(x,x) = B(x,x) + B(x,x)$ vanishes in characteristic 2, we also see that $c_{i,i} = 0$. We can thus write

$$\tilde{B}(x,h) = C(x,h) + C(h,x)$$

where $C(x,h) := \sum_{1 \leqslant i < j \leqslant k} c_{i,j} L_i(x) L_j(h)$ is the lower-triangular component of $\tilde{B}(x,h)$. We then easily verify that $B(x,x) - C(x,x)$ is a linear function of x, and so $P(x)$ can be expressed as the sum of $C(x,x)$ and a linear function, from which the claim $\mathrm{rank}_1(\mathbb{P}) \ll \log_2 1/\delta$ follows. □

We shall establish the following generalisation of the above estimate to higher degree polynomials, provided that the degree does not exceed the characteristic:

Theorem 1.8 (Lack of equidistribution implies bounded rank)**.** *Suppose that an integer d satisfies $0 \leqslant d < |\mathbb{F}|$. Let $\delta \in (0,1]$, and suppose that $P \in \mathcal{P}_d(\mathbb{F}^n)$ is such that $|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))| \geqslant \delta$. Then* $\mathrm{rank}_{d-1}(P) \ll_{\mathbb{F},\delta,d} 1$.

The proof of this theorem is the technical heart of the paper, and will be accomplished in §5. It is possible that the restriction on $|\mathbb{F}|$ can be removed, but our method of proof breaks down when $d \geqslant |\mathbb{F}|$. Certainly the deduction of Theorem 1.4 from Theorem 1.8 breaks down in this case (which of course it must, thanks to Theorem 1.5).

2. Factors and regularity

In this section we give some definitions and results which will be useful in our proof of Theorem 1.8.

Definition 2.1 (Factors and configuration space)**.** *Suppose that $d \geqslant 0$ is an integer and that $M_1, \dots, M_d$ are further non-negative integers. By a* factor *of degree d on $\mathbb{F}^n$ we mean a collection $\mathcal{F} = (P_{i,j})_{1 \leqslant i \leqslant d, 1 \leqslant j \leqslant M_i}$ where $P_{i,j} \in \mathcal{P}_i(\mathbb{F}^n)$ for all i,j. By the* dimension $\dim(\mathcal{F})$ *of $\mathcal{F}$ we mean the quantity $M_1 + \cdots + M_d$. Write $\mathcal{F}_i$ for the i-degree part of $\mathcal{F}$, that is to say the collection $(P_{i,j})_{1 \leqslant j \leqslant M_i}$. Although we are using the term factor to describe nothing more complicated than a collection of polynomials, we encourage the reader to think in addition of the σ-algebra $\sigma(\mathcal{F})$ defined by these polynomials $P_{i,j}$, that is to say the partition of $\mathbb{F}^n$ into* atoms *of the form $\{x : P_{i,j}(x) = c_{i,j}\}$. We write $\Sigma = \mathbb{F}^{M_1} \times \cdots \times \mathbb{F}^{M_d}$ and call this the* configuration space *of $\mathcal{F}$. We write $\Phi : \mathbb{F}^n \to \Sigma$ for the* evaluation map *given by $\Phi(x) = (P_{i,j}(x))_{1 \leqslant i \leqslant d, 1 \leqslant j \leqslant M_i}$.*

We will use the notation of this definition throughout the paper without further comment. Sometimes we will have factors $\mathcal{F}, \mathcal{F}'$ and $\mathcal{F}''$; we will write $P_{i,j}, P'_{i,j}, P''_{i,j}$, $\Sigma, \Sigma', \Sigma''$, M_j, M'_j, M''_j, Φ, Φ', Φ'' and so on for the corresponding polynomials, configuration spaces, dimensions and evaluation maps.

We will frequently need to *extend* a factor into a more *regular* one, by expressing the complicated polynomials in a factor by simpler ones. Our

notation for this concept is as follows. We say that a factor $\mathcal{F}'$ is an *extension* of $\mathcal{F}$ if $\sigma(\mathcal{F}')$ is a (possibly trivial) refinement of $\sigma(\mathcal{F})$. Note that this is *not* the same thing as saying that the collection $(P'_{i,j})$ defining $\mathcal{F}'$ contains the collection $(P_{i,j})$ defining $\mathcal{F}$. For example, the factor defined by the linear polynomials x_1, x_2, x_3 is a refinement of that defined by the polynomials x_1, x_2 and $x_1 + x_2$.

By a *growth function of order d* we mean a non-decreasing function $F : \mathbb{Z}^+ \to \mathbb{R}^+$.

Definition 2.2 (F-regularity)**.** *Let $\mathcal{F}$ be a factor of degree d, and let F be a growth function. We say that $\mathcal{F}$ is* F-regular *if we have*

$$\mathrm{rank}_{i-1}\left(\sum_{j=1}^{M_i} c_{i,j}P_{i,j}\right) \geqslant F(\dim(\mathcal{F}))$$

for all $1 \leqslant i \leqslant d$ and all coefficients $c_{i,1}, \dots, c_{i,M_i} \in \mathbb{F}$ that are not all zero. (In particular, if F is positive, this implies that the polynomials $P_{i,1}, \dots, P_{i,M_i}$ are linearly independent.)

Example 2.3. *If d, F and $M_1, \dots, M_d$ are fixed, and $P_{i,j}$ are chosen uniformly at random from $\mathcal{P}_i(\mathbb{F}^n)$, then the resulting factor $\mathcal{F}$ will be F-regular with probability $1 - o(1)$, where $o(1)$ goes to zero as $n \to \infty$ for fixed $d, F, M_1, \dots, M_d$. Indeed, one should view the polynomials in an F-regular factor as "behaving like" generic polynomials, in that they obey no unexpected algebraic constraints of bounded complexity.*

The following lemma, which allows us to replace take an arbitrary factor $\mathcal{F}$ and find a highly regular extension of it, is absolutely fundamental to our arguments. This generalises [14, Lemma 8.7] to the case of factors of degree 3 or more. The result is faintly analagous in some ways to *Szemerédi's regularity lemma* for graphs and to more recent versions of this for hypergraphs.

Lemma 2.4 (Regularity lemma)**.** *Let $d \geqslant 1$, let F be a growth function, and let $\mathcal{F}$ be a factor of degree d. Then there exists an F-regular extension $\mathcal{F}'$ of $\mathcal{F}$ of degree d satisfying the dimension bound*

$$\dim(\mathcal{F}') \ll_{F,d,\dim(\mathcal{F})} 1.$$

Remark: The actual bound we obtain here, if one worked it out, would have an extremely weak dependence on F, d and $\dim(\mathcal{F})$. Even for quite "reasonable" growth functions F one starts to see functions in the Ackerman hierarchy making an appearance. It is our dependence on this lemma and the rather poor bounds that result from its proof that renders Theorem 1.8 essentially ineffective.

Proof. Fix d and F. We shall induct on the dimension vector $(M_1, \dots, M_d)$ of $\mathcal{F}$ where, of course, $M_i := \dim(\mathcal{F}_i)$. This dimension vector takes values in $\mathbb{Z}_+^d$, which we shall order in reverse lexicographical ordering, that is to say

$(M_1, \dots, M_d) < (M_1', \dots, M_d')$ if there exists $1 \leqslant i \leqslant d$ such that $M_i < M_i'$ and $M_j = M_j'$ for all $i < j \leqslant d$. This turns $\mathbb{Z}_+^d$ into a well-ordered set (with the ordinal type ω^d), and so we can perform strong induction on this space. In other words, we may assume without loss of generality that the claim has already been proven for all smaller dimension vectors.

If $\mathcal{F}$ is already F-regular, then we are done. Otherwise, there exists $i \in [d]$ and a non-trivial linear combination Q_i of the $P_{i,1}, \dots, P_{i,M_i}$ such that $\mathrm{rank}_{i-1}(Q_i) < F(\dim(\mathcal{F}))$, or in other words Q_i is some combination of fewer than $F(\dim(\mathcal{F}))$ polynomials of degree at most $i-1$. By rewriting Q_i in this fashion, we can find an extension $\mathcal{F}''$ of $\mathcal{F}$ with dimension vector

$$(M_1, \dots, M_{i-1} + \lfloor F(\dim(\mathcal{F}))\rfloor, M_i - 1, M_{i+1}, \dots, M_d)$$

(with some obvious modifications in the easy case $i = 1$). Applying the induction hypothesis to $\mathcal{F}''$ we obtain the claim. □

3. A lemma of Bogdanov and Viola

In this section we recall [6, Lemma 25], and provide a proof in the interests of self-containment. This lemma *almost* immediately establishes our main result, Theorem 1.8, except for the presence of some small errors. Our main task in subsequent sections is to eliminate the errors and turn this near-miss result into a proof of Theorem 1.8.

Lemma 3.1 (Bogdanov-Viola lemma)**.** *Let $d \geqslant 0$ be an integer, and let $\delta, \sigma \in (0,1]$ be parameters. Suppose that $P \in \mathcal{P}_d(\mathbb{F}^n)$ is a polynomial of degree d such that*

$$|\mathbb{E}_{x \in \mathbb{F}^n} e_{\mathbb{F}}(P(x))| \geqslant \delta. \tag{3.1}$$

Then there exists a function $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ with $\mathrm{rank}_{d-1}(\tilde{P}) \leqslant |\mathbb{F}|^5/\delta^2\sigma$ such that $\mathbb{P}_{x \in \mathbb{F}^n}(P(x) \neq \tilde{P}(x)) \leqslant \sigma$.

Proof. We remark that the bound on $\mathrm{rank}_{d-1}(\tilde{P})$ is *much* superior to that we will eventually obtain for Theorem 1.8. This is because the Bogdanov-Viola lemma does not rely on the regularity lemma, Lemma 2.4. In fact this bound could even be improved somewhat, but this is not relevent to our work here.

For each $r \in \mathbb{F}$, define a measure $\mu_r : \mathbb{F} \to [0,1]$ by setting

$$\mu_r(t) = \mathbb{P}_{x \in \mathbb{F}^n}(P(x) = t + r)$$

for all $t \in \mathbb{F}$. Then (3.1) implies that $|\sum_{t \in \mathbb{F}} e_{\mathbb{F}}(t)\mu_0(t)| \geqslant \delta$. Noting that

$$\sum_{t \in \mathbb{F}} e_{\mathbb{F}}(t)\mu_0(t) = e_{\mathbb{F}}(d) \sum_t e_{\mathbb{F}}(t)\mu_d(t),$$

we see that

$$\|\mu_0 - \mu_d\| := \sum_t |\mu_0(t) - \mu_d(t)| \geqslant |1 - e_{\mathbb{F}}(d)| \Big| \sum_t e_{\mathbb{F}}(t)\mu_0(t) \Big| \geqslant 4\delta/|\mathbb{F}|$$

if $d \neq 0$, by dint of the inequality $|1 - e^{2\pi i\theta}| \geqslant 4|\theta|$ which holds when $|\theta| \leqslant 1/2$. By translation invariance we conclude that

$$\|\mu_r - \mu_s\| \geqslant 4\delta/|\mathbb{F}| \tag{3.2}$$

whenever $r \neq s$.

Now fix a value of x and let $h \in \mathbb{F}^n$ be chosen at random. Then

$$\mathbb{P}_h(D_hP(x) = t) = \mathbb{P}_h(P(x+h) = t + P(x)) = \mu_{P(x)}(t),$$

that is to say $D_hP(x)$ has the distribution $\mu_{P(x)}$. Now we expect that if a large number $D_{h_1}P(x), \dots, D_{h_k}P(x)$ of points are sampled from this distribution then the observed distribution

$$\mu_{\mathrm{obs}}(h_1, \dots, h_k; x) := \frac{1}{k}\sum_{i=1}^{k} \delta_{D_{h_i}P(x)}$$

should approximate $\mu_{P(x)}$. In view of the separation property (3.2), this ought to give us a good chance of recovering $P(x)$.

Choose $k \geqslant \frac{|\mathbb{F}|^5}{2\sigma\delta^2}$, and sample $h_1, \dots, h_k$ independently at random from $\mathbb{F}^n$. Motivated by the above discussion, we define $\tilde{P}_{h_1,\dots,h_k}(x)$ to be that value of $r \in \mathbb{F}$ for which $\|\mu_{\mathrm{obs}}(h_1, \dots, h_k; x) - \mu_r\|$ is minimal. Note that $\tilde{P}_{h_1,\dots,h_k}$ is measurable with respect to the set of functions $D_{h_1}P(x), \dots, D_{h_k}P(x)$, each of which is a polynomial of degree at most $d-1$. Thus

$$\mathrm{rank}_{d-1}(\tilde{P}_{h_1,\dots,h_k}) \leqslant k.$$

It remains to show that, at least for some choice of $h_1, \dots, h_k$, the function $\tilde{P}_{h_1,\dots,h_k}$ approximates P. Now if $\tilde{P}_{h_1,\dots,h_k}(x) \neq P(x)$ then it follows from the separation property (3.2) that

$$\|\mu_{\mathrm{obs}}(h_1, \dots, h_k, x) - \mu_{P(x)}\| \geqslant 2\delta/|\mathbb{F}|.$$

We claim that for fixed x the probability of this happening (over random choices of $h_1, \dots, h_k$) is at most σ. Summing over x, it then follows that there is at least one choice of $h_1, \dots, h_k$ for which

$$\#\{x : P(x) \neq \tilde{P}_{h_1,\dots,h_k}(x)\} \leqslant \sigma|\mathbb{F}^n|,$$

and the lemma follows upon taking $\tilde{P} := \tilde{P}_{h_1,\dots,h_k}$.

Fix $x \in \mathbb{F}^n$ and a value of $t \in \mathbb{F}$, and write $Y_i = 1_{D_{h_i}P(x)=t}$. To establish the claim, it suffices to show that

$$\mathbb{P}\left(\left|\frac{Y_1 + \cdots + Y_k}{k} - \mu_{P(x)}(t)\right| \geqslant \frac{2\delta}{|\mathbb{F}|}\right) \leqslant \frac{\sigma}{|\mathbb{F}|}.$$

Noting that the Y_i are i.i.d. Bernouilli random variables with means $\overline{Y} = \mu_{P(x)}(t)$, this follows from a suitable version of the law of large numbers. In this case we may use the inequality

$$\mathbb{P}\left(\left|\frac{Y_1 + \cdots + Y_k}{k} - \overline{Y}\right| \geqslant \eta\right) \leqslant \frac{1}{4k\eta^2},$$

which follows from Chebyshev's inequality. □

Remark: When $|\mathbb{F}| = 2$, the above proof has a pleasant interpretation. The value of $\tilde{P}_{h_1,\dots,h_k}(x)$ is then obtained by "majority vote" amongst the values of $D_{h_i}P(x)$.

4. Counting lemmas

We shall prove Theorem 1.8 by induction. Accordingly, we begin by first describing some *consequences* of Theorem 1.8 at a given degree d, which are already of some independent interest. These consequences complement the regularity lemma in much the same way that "counting lemmas" in graph theory complement the Szemerédi regularity lemma.

Lemma 4.1 (Size of atoms)**.** *Let $d \geqslant 1$, and $\varepsilon > 0$. Suppose that Theorem 1.8 is true for degrees up to d. Then there exists a growth function F (depending on d and ε) such that if $\mathcal{F}$ is an F-regular factor of degree d on $\mathbb{F}^n$ then we have the estimate*

$$\mathbb{P}_{x\in\mathbb{F}^n}(\Phi(x) = t) = (1 + O(\varepsilon))\frac{1}{|\Sigma|} \tag{4.1}$$

for all configurations $t \in \Sigma$. In words, all the atoms in the σ-algebra $\sigma(\mathcal{F})$ have roughly the same size.

Remark: Recall that $\Sigma = \mathbb{F}^{M_1} \times \cdots \times \mathbb{F}^{M_d}$ is the configuration space associaed to the factor $\mathcal{F}$, and that $\Phi : \mathbb{F}^n \to \Sigma$ is the evaluation map.

Proof. We may expand the condition $\Phi(x) = t$ using Fourier analysis on Σ to obtain

$$\mathbb{P}_x(\Phi(x) = t) = \frac{1}{|\Sigma|}\sum_{r\in\Sigma}\mathbb{E}_{x\in\mathbb{F}^n} e_{\mathbb{F}}(r\cdot(\Phi(x) - t)).$$

It therefore suffices to show that

$$\mathbb{E}_{x\in\mathbb{F}^n} e_{\mathbb{F}}\left(\sum_{i=1}^{d} Q_i\right) = O\left(\varepsilon/|\mathbb{F}|^{\dim(\Sigma)}\right) \tag{4.2}$$

whenever the $Q_i \in \mathrm{Span}(\mathcal{F}_i)$ are not all zero. Let $s \in [d]$ be the largest integer for which Q_s is non-zero. As $\mathcal{F}$ is F-regular, we have $\mathrm{rank}_{s-1}(Q_s) \geqslant F(\dim(\mathcal{F}))$. On the other hand, $\sum_{i=1}^{d} Q_i$ differs from Q_s by an element of $\mathcal{P}_{s-1}(V)$. Thus

$$\mathrm{rank}_{s-1}\left(\sum_{i=1}^{d} Q_i\right) \geqslant F(\dim(\mathcal{F})) - 1.$$

If we choose F to sufficiently rapidly growing depending on ε and d, we can thus invoke Theorem 1.8 to obtain (4.2) as required. □

In addition to understanding the distribution of $\Phi(x)$, it turns out to be important to have an understanding of how k-dimensional *parallelepipeds* are distributed in configuration space. That is, we study the distribution of $(\Phi(x + \omega\cdot h))_{\omega\in\{0,1\}^k}$ in $\Sigma^{\{0,1\}^k}$, where $h = (h_1, \dots, h_k)$ is a k-tuple of

elements of $\mathbb{F}^n$. When $k = 2$, for example, we are interested in the 4-tuple $(\Phi(x), \Phi(x+h_1), \Phi(x+h_2), \Phi(x+h_1+h_2))$. We prepare the ground for this study with some definitions.

Definition 4.3 (Faces and lower faces)**.** *Let $k \geqslant 1$ be an integer and suppose that $0 \leqslant k' \leqslant k$. A subset $F \subseteq \{0,1\}^k$ is called a* face *of dimension k' if it has the form*

$$F = \{\omega \in \{0,1\}^k : \omega_i = \delta_i \quad \text{for } i \in I\},$$

where $I \subseteq [k]$ has size $k - k'$ and each δ_i is either 0 or 1. If all of the δ_i are zero then we say that F is a lower face*. A lower face of dimension k' can be identified with the power set of $[k] \setminus I$, which is a set of size k'.*

Suppose that we have a parallelepiped $(x + \omega \cdot h)_{\omega \in \{0,1\}^k}$ in $\mathbb{F}^n$, where $h = (h_1, \dots, h_k)$ is a k-tuple of elements of $\mathbb{F}^n$. Consider the image $(\Phi(x + \omega \cdot h))_{\omega \in \{0,1\}^k} \in \Sigma^{\{0,1\}^k}$. This cannot be arbitrary: indeed we have the "obvious" constraints coming from the relations

$$\sum_{\omega \in F} (-1)^{|\omega|} P_{i,j}(x + \omega \cdot h) = 0$$

whenever $F \subseteq \{0,1\}^k$ is a face of dimension at least $i + 1$, and $|\omega| := \omega_1 + \cdots + \omega_k$. To model these obvious constraints, we introduce some more notation.

Definition 4.4 (Face vectors and parallelepiped constraints)**.** *Suppose that $i_0 \in [d]$, that $j_0 \in [M_{i_0}]$ and that $F \subseteq \{0,1\}^k$. Consider the vector $r(i_0, j_0, F) \in \Sigma^{\{0,1\}^k}$ for which $r_{i,j}(\omega) = (-1)^{|\omega|}$ if $i = i_0$, $j = j_0$ and $\omega \in F$, and is zero otherwise. We call such a vector a* face vector*. If F is a lower face then we speak of a* lower face vector*. If $\dim(F) \geqslant i_0 + 1$ we say that the face vector (or lower face vector) is* relevant*. We say that $(t(\omega))_{\omega \in \{0,1\}^k} \in \Sigma^{\{0,1\}^k}$* satisfies the parallelepiped constraints *if it is orthogonal to all the relevant lower face vectors.*

Remarks: The motivation for this definition, of course, is that for any x, h_1, $\dots$, h_k the vector $(\Phi(x+\omega\cdot h))_{\omega\in\{0,1\}^k} \in \Sigma^{\{0,1\}^k}$ satisfies the parallelepiped constraints. At first sight the fact that we have restricted attention to *lower* face vectors may look curious. However it turns out (and is not hard to prove) that the set of relevant face vectors in $\Sigma^{\{0,1\}^k}$ is spanned by the relevant lower face vectors. We will not require this fact.

Write $\Sigma_\square \subseteq \Sigma^{\{0,1\}^k}$ for the subspace of vectors in $\Sigma^{\{0,1\}^k}$ satisfying the parallelepiped constraints.

Lemma 4.5 (Dimension of $\Sigma_\square$)**.** *Suppose that $k > d$. Then we have*

$$\dim(\Sigma_\square) = \sum_{i=1}^{d} M_i \sum_{0 \leqslant j \leqslant i} \binom{k}{j}.$$

Proof. Since $\dim(\Sigma^{\{0,1\}^k}) = 2^k(M_1 + \cdots + M_d) = \sum_{i=1}^d M_i \sum_j \binom{k}{j}$, it suffices to show that the dimension of the space spanned by the relevant lower face vectors is $\sum_{i=1}^d M_i \sum_{j>i} \binom{k}{j}$. This is precisely the number of different relevant lower face vectors, and so we must only show that the lower face vectors are linearly independent. To do this, we may clearly work with a fixed choice of i and j, since the supports of the face vectors $r(i,j,F)$ are disjoint for different pairs (i,j). Suppose there is some linear relation

$$\sum_F a_F r(i,j,F) = 0.$$

Among all lower faces F for which $a_F \neq 0$, suppose that F_0 contains the largest element ω_0 in the lexicographic order on $\{0,1\}^k$. Comparing coefficients of ω_0 we see that $a_{F_0} = 0$, contrary to assumption. □

If the factor $\mathcal{F}$ is F-regular for some sufficiently rapid growth function F, it turns out that the parallelepiped constraints we have written down are the only relevant ones in a rather strong sense.

Proposition 4.6 (Counting parallelepipeds)**.** *Suppose that $|\mathbb{F}|, k > d$, and suppose that Theorem 1.8 is true for degrees up to d. Let $\varepsilon \in (0,1)$ be a parameter and suppose that F grows sufficiently quickly (depending on k, d and ε). Suppose that the factor $\mathcal{F}$ has degree at most d and is F-regular. Suppose that $t_\square \in \Sigma_\square$, and that $x \in \mathbb{F}^n$ is a point with $\Phi(x) = t_\square(0)$. Then the number of $h \in (\mathbb{F}^n)^k$ such that $\Phi(x + \omega \cdot h) = t_\square(\omega)$ for all $\omega \in \{0,1\}^k$ is $1 + O_k(\varepsilon)$ times $|\mathbb{F}|$ to the power $nk - \sum_{i=1}^d M_i \sum_{1 \leqslant j \leqslant i} \binom{k}{j}$.*

Remark: Note carefully that we have been able to fix the basepoint x; this is important in applications of the proposition. This is why j now only ranges from 1 to i rather than from 0 to i as in Lemma 4.5.

Proof. Write $\Phi_\square(h)$ for the vector $(\Phi(x + \omega \cdot h))_{\omega \in \{0,1\}^k}$ in $\Sigma^{\{0,1\}^k}$. We seek the number of h for which $\Phi_\square(h) = t_\square$; by harmonic analysis on $\Sigma^{\{0,1\}^k}$ this may be expanded as

$$|\mathbb{F}|^{nk} \left|\Sigma^{\{0,1\}^k}\right|^{-1} \sum_{r_\square \in \Sigma^{\{0,1\}^k}} \mathbb{E}_{h \in (\mathbb{F}^n)^k} e_{\mathbb{F}}(r_\square \cdot (\Phi_\square(h) - t_\square)). \tag{4.3}$$

Now when $r_\square$ lies in the space W spanned by the relevant lower face vectors together with the vectors $r(i,j,0)$ we have $r_\square \cdot (\Phi_\square(h) - t_\square) = 0$, since both $\Phi_\square(h)$ and $t_\square$ satisfy the parallelepiped constraints and $\Phi_\square(h)(0) = t_\square(0)$. Since the lower face vectors are linearly independent the contribution from these $r_\square$ to the sum (4.3) is $|\mathbb{F}|$ to the power $nk - \sum_{i=1}^d M_i \sum_{1 \leqslant j \leqslant i} \binom{k}{j}$. To conclude the argument it certainly suffices to show that the contribution from each $r_\square \notin W$ is small in the sense that

$$|\mathbb{E}_{h \in (\mathbb{F}^n)^k} e_{\mathbb{F}}(r_\square \cdot \Phi_\square(h))| \leqslant \varepsilon |\mathbb{F}|^{-2^k \dim(\Sigma)}. \tag{4.4}$$

Such an exponential sum is unaltered in magnitude if an arbitrary element of W is added to $r_\square$. By repeated operations of this type, directed so as to reduce the largest element in the ω-support of each $(r_\square(\omega))_{i,j}$ in the lexicographic order on $\{0,1\}^k$, we may assume that $(r_\square(\omega))_{i,j} = 0$ unless $|\omega| \leqslant i$. Since $r_\square$ is not in W, there is at least one choice of i, j and at least one $\omega \neq 0$ for which $(r_\square(\omega))_{i,j} \neq 0$. Amongst all such triples (i, j, ω), choose one with the largest value of i, say $i = i_0$. For this value of $i = i_0$ choose (j_0, ω_0) with $s = |\omega_0|$ maximal, still subject to the condition that $(r_\square(\omega_0))_{i_0,j_0} \neq 0$. Note that $1 \leqslant s \leqslant i$. By relabelling the cube $\{0,1\}^k$ we may assume that $\omega_0 = 1^s 0^{k-s}$. By construction, any triple (i, j, ω) satisfies one of the following properties:

(1) $i > i_0$ and $\omega = 0$;
(2) $i = i_0$ and $\omega = \omega_0$;
(3) $i = i_0$ and at least one of the coordinates ω_l, $1 \leqslant l \leqslant s$, is zero;
(4) $i < i_0$.

Since $1 \leqslant s \leqslant i \leqslant k$, the sum in (4.4) may then be written as an average (over $h_{s+1}, \dots, h_k$) of sums of the form

$$\mathbb{E}_{h_1,\dots,h_s} e_{\mathbb{F}}(P(x + h_1 + \cdots + h_s) + Q(h_1, \dots, h_s)),$$

where P is not zero and lies in $\mathrm{Span}(\mathcal{F}_i)$, and Q has degree at most $s-1$ as a polynomial in $h_1, \dots, h_s$. Such a sum may be written as

$$\mathbb{E}_{h_1,\dots,h_s} \mathbf{b}_1(h_2, \dots, h_s) \dots \mathbf{b}_s(h_1, \dots, h_{s-1}) e_{\mathbb{F}}(P(x + h_1 + \cdots + h_s)),$$

where each $\mathbf{b}$ is a bounded function which does not depend on h_i. By introducing dummy variables we may assume that $s = i$. Applying the Cauchy-Schwarz inequality i times to eliminate the bounded functions $\mathbf{b}$, we see that the sum in (4.4) may be bounded thus:

$$|\mathbb{E}_{h \in (\mathbb{F}^k)^n} e_{\mathbb{F}}(r_\square \cdot \Phi_\square(h))| \leqslant \left(\mathbb{E}_{h_1,\dots,h_i} e_{\mathbb{F}}(D_{h_1} \cdots D_{h_i} P(\cdot))\right)^{1/2^i}.$$

Note that this derivative is, for fixed $h_1, \dots, h_i$, simply a constant; we write it as $\partial^i P(h_1, \dots, h_i)$. It follows that if (4.4) is false then

$$|\mathbb{E}_{h_1,\dots,h_i} e_{\mathbb{F}}(\partial^i P(h_1, \dots, h_i))| \geqslant (\varepsilon |\mathbb{F}|^{-2^k \dim(\Sigma)})^{2^i}.$$

Applying Theorem 1.8 at degree $i \leqslant d$ and with $V = (\mathbb{F}^n)^i$ we see that

$$\mathrm{rank}_{i-1}(\partial^i P) \ll_{k,\varepsilon,\dim(\Sigma)} 1.$$

Note however that we have the Taylor expansion

$$P(x) = \frac{1}{i!} \partial^i P(x, \dots, x) + Q(x)$$

for some polynomial Q of degree at most $i-1$ (this is the only point in the whole paper where we use the assumption that $|\mathbb{F}| > d \geqslant i$, in order to ensure invertibility of $i!$). It follows that

$$\mathrm{rank}_{i-1}(P) \ll_{k,\varepsilon,\dim(\Sigma)} 1.$$

This contradicts the F-regularity of the factor $\mathcal{F}$ if F is assumed to grow sufficiently rapidly. □

5. Proof of Theorem 1.8

In this section we complete the proof of Theorem 1.8. Our starting point is the lemma of Bogdanov and Viola, stated as Lemma 3.1 in this paper. We urge the reader to recall the statement now. In view of that lemma, it suffices to establish the following proposition.

Proposition 5.1 (Polynomials which are almost low-rank are low-rank)**.** *Suppose that $d \geqslant 1$ is an integer, and that Theorem 1.8 holds for all degrees up to $d-1$. Let $\sigma_d > 0$ be a small quantity to be specified later. Suppose that $P \in \mathcal{P}_d(\mathbb{F}^n)$ and that $\mathcal{F}$ is an F-regular factor of degree $d-1$. for some growth function which grows suitably rapidly in terms of d. Suppose that $\tilde{P} : \mathbb{F}^n \to \mathbb{F}$ is an $\mathcal{F}$-measurable function and that $\mathbb{P}(P(x) = \tilde{P}(x)) \geqslant 1 - \sigma_d$. Then P is itself $\mathcal{F}$-measurable.*

Proof of Theorem 1.8 assuming Proposition 5.1. This is almost immediate. By induction we may fix $d \geqslant 1$ and assume that Theorem 1.8 holds for all degrees up to $d-1$. Take the function $\tilde{P}$ appearing in the conclusion of Lemma 3.1. By construction, $\tilde{P}$ is measurable with respect to some factor $\mathcal{F}_0$ of degree at most $d-1$ and dimension no more than $|\mathbb{F}|^5/\delta^2\sigma$. By Lemma 2.4 we may extend $\mathcal{F}_0$ to a factor $\mathcal{F}$ which is F-regular and satisfies $\dim(\mathcal{F}) \ll_{F,d,\delta,\mathbb{F}} 1$. The function $\tilde{P}$ is manifestly $\mathcal{F}$-measurable, and so the result follows upon applying Proposition 5.1. □

Proof of Proposition 5.1. We use the same notation for the factor $\mathcal{F}$ that was introduced in Definition 2.1. In particular this factor is defined by polynomials $P_{i,j} \in \mathcal{P}_i(\mathbb{F}^n)$: these should not be confused with the polynomial P which is the subject of Proposition 5.1.

For the purposes of an initial discussion write X for the set of points in $\mathbb{F}^n$ for which $P(x) = \tilde{P}(x)$, thus $|X| \geqslant (1-\sigma_d)|\mathbb{F}^n|$. The key idea is that we may use $(d+1)$-dimensional parallelepipeds in X to create new points x' for which $P(x')$ does not depend on which atom of $\mathcal{F}$ the point x' lies in. There are two procedures we might use:

1. *Completing atoms.*
   Suppose that $x, h_1, \dots, h_{d+1}$ are such that all 2^{d+1} points $x + \omega \cdot h$ lie in the same atom A of $\sigma(\mathcal{F})$. Suppose in addition that $x + \omega \cdot h \in X$ whenever $\omega \neq 0$. Then using the relation $\sum_\omega (-1)^{|\omega|} P(x + \omega \cdot h) = 0$ and the fact that $\tilde{P}$ is constant on A, we see that x also lies in X.

2. *Creating new atoms on which P is constant.*
   Suppose that A is an atom of $\sigma(\mathcal{F})$ such that there are atoms A_ω, $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$ with the following property. For any $x \in A$, there are $h_1, \dots, h_{d+1} \in \mathbb{F}^n$ such that $x + \omega \cdot h \in A_\omega$ for all $\omega \in \{0,1\}^{d+1} \setminus 0$. Then

if P is constant on each of the A_ω, it is also constant on A. This follows from the relation $\sum_\omega (-1)^{|\omega|} P(x+\omega\cdot h) = 0$ once again.

It is in fact possible to perform Procedures 1 and 2 simultaneously, but the exposition is fractionally clearer if the urge to do this is suppressed.

Let us start with an analysis of Procedure 1. It is easy to see using Lemma 4.1 that for $1 - O(\sqrt{\sigma_d})$ of the atoms in $\mathcal{B}$ we have $P_{x\in A}(P(x) = \tilde{P}(x)) \geqslant 1 - O(\sqrt{\sigma_d})$. We say that P is *almost constant* on such atoms, and our task is to show that P is actually 100% constant on each such atom.

Suppose that P is almost constant on the atom $A = \Phi^{-1}(t)$, and write $A' \subseteq A$ for the set where $P = \tilde{P}$.

Lemma 5.2 (Avoiding bad parallelepipeds)**.** *Let the notation and assumptions be as above. Suppose that σ_d is chosen sufficiently small. Fix an $x \in A$. Then there is h so that all of the vertices $x + \omega\cdot h$, $\omega \neq 0^{d+1}$, lie in A'.*

Proof. Let $N_\square(x)$ denote the number of parallelepipeds $(x+\omega\cdot h)_{\omega\in\{0,1\}^{d+1}}$, all of whose vertices lie in A. The vector $(t,t,\dots,t) \in \Sigma^{\{0,1\}^{d+1}}$ trivially satisfies the parallelepiped constraints, and so by Proposition 4.6 we have

$$N_\square(x) \sim |\mathbb{F}|^{n(d+1) - \sum_{i=1}^d M_i \sum_{1\leqslant j\leqslant i} \binom{d+1}{j}} \tag{5.1}$$

if F is sufficiently rapidly growing.

The number $N_\square(x)$ of parallelepipeds in A is thus quite large. Unfortunately, this does not immediately imply that the number of paralleopipeds in A' is large, as the $N_\square(x)$ parallelepipeds in A may all be intersecting the small set $A\backslash A'$. However, it will turn out that such a concentration in $A\backslash A'$ can be picked up via the Cauchy-Schwarz inequality, as it will force into existence an anomalously large number of *pairs* of parallelepipeds that share an additional vertex in common besides x. The main difficulty in the proof then lies in counting number of such pairs properly.

We turn to the details. It suffices to show, for each fixed $\omega_0 \in \{0,1\}^{d+1} \setminus 0^{d+1}$, that the number of parallelepipeds $(x+\omega\cdot h)_{\omega\in\{0,1\}^{d+1}}$, all of whose vertices lie in A, and with $x + \omega_0\cdot h \in A \setminus A'$, is less than $2^{-d-2} N_\square(x)$. The number of such "bad" parallelepipeds may be written as

$$\sum_u 1_{A\setminus A'}(u) \sum_h 1_{x+\omega_0\cdot h = u},$$

and we may use the Cauchy-Schwarz inequality to bound this above by

$$|A\setminus A'|^{1/2} \big| \{(h,h') : x+\omega\cdot h, x+\omega'\cdot h' \in A \text{ for all } \omega,\omega' \in \{0,1\}^{d+1}, x+\omega_0\cdot h = x+\omega_0\cdot h'\}\big|^{1/2}.$$

Thus if σ_d is chosen so small that $|A \setminus A'| \leqslant 2^{-2d-5}|A|$, it suffices to show that

$$\big|\{(h,h') : x+\omega\cdot h, x+\omega'\cdot h' \in A \text{ for all } \omega,\omega' \in \{0,1\}^{d+1}, x+\omega_0\cdot h = x+\omega_0\cdot h'\}\big| \leqslant \frac{N_\square(x)^2}{|A|}(1+O(\varepsilon))$$

for some sufficiently small $\varepsilon > 0$.

By relabelling the cube $\{0,1\}^{d+1}$ if necessary, this may be recast as the problem of counting the number of $h, h' \in (\mathbb{F}^n)^{d+1}$ satisfying the constraint

$$h_1 + \cdots + h_s = h'_1 + \cdots + h'_s$$

and for which the two parallelepipeds

$$\square_1 := (x+\omega\cdot h)_{\omega\in\{0,1\}^{d+1}}$$

and

$$\square_2 := (x+\omega\cdot h')_{\omega\in\{0,1\}^{d+1}}$$

lie in A. Substituting (5.1) and the approximate size of $|A|$ (cf. Lemma 4.1) into (5), we see that our task is to establish that the number of such h, h' is at most $1+O(\varepsilon)$ times $|\mathbb{F}|$ to the power $n(2d+1)+\sum_{i=1}^d M_i(1-2\sum_{1\leqslant j\leqslant i}\binom{d+1}{j})$.

The parallelepipeds $\square_1$ and $\square_2$ share the common vertices x and $x+h_1+\cdots+h_s$. Note that $\square_1$ and $\square_2$ may be embedded inside a $(2d+1)$-dimensional parallelepiped

$$\tilde{\square} := (x+\omega\cdot y)_{\omega\in\{0,1\}^{2d+1}},$$

where

$$y := (h_1,\ldots,h_{s-1},h_s-h'_1-\cdots-h'_{s-1},h_{s+1},\ldots,h_{d+1},h'_1,\ldots,h'_{s-1},h'_{s+1},\ldots,h'_{d+1}).$$

Thus, writing $\square_1$ corresponds to the indices

$$\omega \in \{0,1\}^{d+1}\cdot(e_1,\ldots,e_{s-1},e_s+e_{d+2}+\cdots+e_{d+s},e_{s+1},\ldots,e_{d+1}), \tag{5.2}$$

and $\square_2$ to the indices

$$\omega \in \{0,1\}^{d+1}\cdot(e_{d+2},\ldots,e_{d+s},e_1+\cdots+e_s,e_{d+s+1},\ldots,e_{2d+1}), \tag{5.3}$$

where we use the usual dot product

$$(\omega_1,\ldots,\omega_{d+1})\cdot(v_1,\ldots,v_{d+1}) := \omega_1v_1+\cdots+\omega_{d+1}v_{d+1}.$$

Suppose that $i \in [d]$ and $j \in [M_i]$. Then $P_{i,j}(x+\omega\cdot y)$ is a polynomial of total degree at most i in $\omega_1,\ldots,\omega_{2d+1}$. Using the fact that $\omega = \omega^2 = \omega^3 = \ldots$ for $\omega \in \{0,1\}$, we see that there exists a polynomial $Q_{i,j} : \mathbb{Z}^{2d+1} \to \mathbb{F}$ with total degree at most i and degree at most 1 in each of $\omega_1,\ldots,\omega_{2d+1}$ with the property that

$$P_{i,j}(x+\omega\cdot y) = Q_{i,j}(\omega)$$

for $\omega \in \{0,1\}^{2d+1}$. In fact this extension is unique, as the following lemma shows.

Lemma 5.3 (Extension lemma)**.** *Suppose that $Q : \mathbb{Z}^k \to \mathbb{F}$ is a polynomial in variables $x_1, \dots, x_k$ of total degree with degree at most one in each x_j. Suppose that $Q(x_1, \dots, x_k)$ is equal to zero for $(x_1, \dots, x_k) \in \{0,1\}^k$. Then $Q \equiv 0$ identically.*

Proof. This appears, for example, as [1, Lemma 2.1]. We proceed by induction on k, the result being trivial when $k = 1$. We may write

$$Q(x_1, \dots, x_k) = R(x_1, \dots, x_{k-1}) + x_k S(x_1, \dots, x_{k-1}),$$

where both R and S have degree at most one in each x_j. Noting that

$$R(x_1, \dots, x_{k-1}) = Q(x_1, \dots, x_{k-1}, 0)$$

and that

$$S(x_1, \dots, x_{k-1}) = Q(x_1, \dots, x_{k-1}, 1) - Q(x_1, \dots, x_{k-1}, 0),$$

we see that $R(x_1, \dots, x_{k-1}) = S(x_1, \dots, x_{k-1}) = 0$ for all $x_j \in \{0,1\}$. By the inductive hypothesis this implies that $R \equiv S \equiv 0$ identically. □

It follows from Lemma 5.3, (5.2), (5.3) and the fact that $P_{i,j}(\square_1)$ and $P_{i,j}(\square_2)$ are fixed that $Q_{i,j}(\omega)$ is fixed for ω in both of the $d+1$-dimensional lattices

$$\Lambda := \mathbb{Z}^{d+1} \cdot (e_1, \dots, e_{s-1}, v, e_{s+1}, \dots, e_{d+1})$$

and

$$\Lambda' := \mathbb{Z}^{d+1} \cdot (e_{d+2}, \dots, e_{d+s}, v, e_{d+s+1}, \dots, e_{2d+1}),$$

where $v \in \mathbb{Z}^{2d+1}$ is the vector

$$v := e_1 + \dots + e_s + e_{d+2} + \dots + e_{d+s}.$$

A second application of Lemma 5.3, noting that $2d > i$, confirms that $Q_{i,j}$ is determined on

$$\mathbb{Z}^{2d} \cdot (e_1, \dots, e_{s-1}, e_{s+1}, \dots, e_{2d+1}) + \{0,1\} \cdot v$$

by its values on

$$S := \{0,1\} \cdot (e_1, \dots, e_{s-1}, e_{s+1}, \dots, e_{2d+1}, v).$$

In particular we see that $Q_{i,j}(\omega)$, and hence $P_{i,j}(x + \omega \cdot y)$, is determined for $\omega \in \{0,1\}^{2d+1}$ by its values on S. Since $Q_{i,j}$ has degree at most i we see that it is determined on S by its values at arguments which are the sum of at most i elements from $\{e_1, \dots, e_{s-1}, e_{s+1}, \dots, e_{2d+1}, v\}$.

Of the $\sum_{0 \leqslant j \leqslant i} \binom{2d+1}{j}$ possible choices for the values of the polynomials $Q_{i,j}$ at these arguments, $2\sum_{0 \leqslant j \leqslant i} \binom{d+1}{j} - 2$ of them are already fixed for us since $Q_{i,j}$ is fixed in both Λ and Λ'. It follows that the number of choices of $(P_{i,j}(x + \omega \cdot y))_{\omega \in \{0,1\}^{2d+1}}$ is at most $|\mathbb{F}|$ to the power $1 + \sum_{1 \leqslant j \leqslant i} \binom{2d+1}{j} - 2\binom{d+1}{j}$. Summing over i and j, it follows that the number of choices for $\Phi(\tilde{\square})$ subject to our constraints on $\Phi(\square_1)$ and $\Phi(\square_2)$ is at most $|\mathbb{F}|$ to the power $\sum_{i=1}^{d} M_i\big(1 + \sum_{1 \leqslant j \leqslant i} \binom{2d+1}{j} - 2\binom{d+1}{j}\big)$.

For each such choice the number of $\tilde{\square}$ is, by Proposition 4.6, $1 + O(\varepsilon)$ times $|\mathbb{F}|$ to the power $n(2d+1) - \sum_{i=1}^{d-1} M_i \sum_{1 \leqslant j \leqslant i} \binom{2d+1}{j}$, and so the total number of $\tilde{\square}$ is $1 + O(\varepsilon)$ times $|\mathbb{F}|$ to the power

$$n(2d+1) + \sum_{i=1}^{d-1} M_i \left(1 - 2 \sum_{1 \leqslant j \leqslant i} \binom{d+1}{j}\right),$$

which is what we wanted to prove. This concludes the proof of Lemma 5.2. □

Recall that $A' \subseteq A$ is the set of points where $P(x) = \tilde{P}(x)$. Now A is an atom in the factor $\mathcal{F}$, which has degree $d-1$, and P is a polynomial of degree d. We therefore see that if all the points $x + \omega \cdot h$, $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$, lie in A' then so does x. It follows from Lemma 5.2 that $A' = A$.

This completes the analysis of Procedure 1, and we find ourselves in the situation that $P(x) = \tilde{P}(x)$ on $1 - O(\sqrt{\sigma_d})$ of the atoms in $\sigma(\mathcal{F})$. Call these the *good* atoms. To perform procedure 2, we need only show that for any (bad) atom $A = A_0$ there are good atoms A_ω, $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$, such that the sequence of coordinates $t_\square = \Phi(A_\omega) \in \Sigma^{\{0,1\}^{d+1}}$ satisfies the parallelepiped constraints. To do this it suffices to find just a single parallelepiped $(x + \omega \cdot h)_{\omega \in \{0,1\}^d}$ for which all of $x + \omega \cdot h$, $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$, lie in good atoms. To see that this is possible, fix $x \in A_0$ and pick $h_1, \dots, h_{d+1}$ at random. It is clear that for any fixed $\omega \neq 0^{d+1}$, the probability that $x + \omega \cdot h$ lies in a good atom is the same as the probability that a random element of $\mathbb{F}^n$ lies in a good atom, which is $1 - O(\sqrt{\sigma_d})$ by Lemma 4.1. If $\sigma_d \leqslant c 2^{-2d}$ for sufficiently small c it follows that there is indeed positive probability that all of the $x + \omega \cdot h$, $\omega \in \{0,1\}^{d+1} \setminus 0^{d+1}$, lie in good atoms.

We have now successfully performed Procedures 1 and 2. By earlier remarks, this concludes the proof of Proposition 5.1 and hence, by the remarks at the start of the section, that of Theorem 1.8. □

6. Inverse theorems for the Gowers norm

We can now give a fairly quick proof of Theorem 1.4. We begin with a preliminary result which is already of interest.

Proposition 6.1. *Suppose that $|\mathbb{F}| > d+1 \geqslant 2$ and that $\delta > 0$, let $P \in \mathcal{P}_{d+1}(\mathbb{F}^n)$, and write $f(x) := e_{\mathbb{F}}(P(x))$. Suppose that $\|f\|_{U^{d+1}} \geqslant \delta$. Then $\mathrm{rank}_d(P) \ll_{d,\delta} 1$.*

Proof. Write $\partial^{d+1} P(h_1, \dots, h_{d+1}) := D_{h_1} \cdots D_{h_{d+1}} P(x)$. Since P has degree $d+1$, this does not depend on x. From the definition of the U^{d+1} norm, we have

$$|\mathbb{E}_{h \in (\mathbb{F}^n)^{d+1}} e_{\mathbb{F}}(\partial^{d+1} P(h))| = \|f\|_{U^{d+1}}^{2^{d+1}} \geqslant \delta^{2^{d+1}}.$$

Applying Theorem 1.8, we conclude that

$$\mathrm{rank}_d(\partial^{d+1} P) \ll_{d,\delta} 1.$$

But since $|\mathbb{F}| > d+1$ we have the Taylor expansion

$$P(x) = \frac{1}{(d+1)!}\partial^{d+1}P(x,x,\dots,x) + Q(x),$$

where $\deg Q \leqslant d$. Thus the rank of P is itself bounded by $O_{d,\delta}(1)$, as required. □

Proof of Theorem 1.4. We fix d and induct on k. The cases $k \leqslant d$ are trivial (since $\|f\|_{u^{d+1}} = 1$ in these cases), so we first verify the case $k = d+1$. In this case, we know from Proposition 6.1 that $\mathrm{rank}_d(P) \ll_{d,\delta} 1$, thus we can express $f(x) = e_{\mathbb{F}}(P(x))$ as some function of $O_{d,\delta}(1)$ polynomials of degree at most d. By Fourier analysis, we can therefore obtain a representation

$$f(x) = \sum_{j=1}^{J} c_j e_{\mathbb{F}}(Q_j(x)),$$

where $J = O_{d,\delta}(1)$, $Q_j \in \mathcal{P}_d(\mathbb{F}^n)$, and c_j are complex numbers of magnitude $O_{d,\delta}(1)$ for all $j \in [J]$. It follows immediately that f has inner product at $\gg_{d,\delta} 1$ with at least one of the functions $e_{\mathbb{F}}(Q_i(x))$, and therefore $\|f\|_{u^{d+1}} \gg_{d,\delta} 1$ as desired.

Now suppose that $k > d$ and the claim has already been proven for polynomials of degree k. Suppose that $P \in \mathcal{P}_{k+1}(\mathbb{F}^n)$, that $f(x) := e_{\mathbb{F}}(P(x))$ and that $\|f\|_{U^{d+1}} \geqslant \delta$. By the monotonicity of Gowers norms (see e.g. [19, Chapter 11]) we have

$$\|f\|_{U^{k+1}} \geqslant \delta$$

and thus by Proposition 6.1 we obtain

$$\mathrm{rank}_k(P) \ll_{k,\delta} 1.$$

Let F be a growth function (depending on k, δ, d) to be chosen later. Applying Lemma 2.4, we can find an F-regular factor $\mathcal{F}$ of degree k and dimension $O_{F,k,d,\delta}(1)$ such that P is measurable with respect to $\sigma(\mathcal{F})$. By Fourier expansion, we can thus express

$$f(x) = \sum_{Q_1 \in \mathrm{Span}(\mathcal{F}_1),\dots,Q_k \in \mathrm{Span}(\mathcal{F}_k)} c_{Q_1,\dots,Q_k} e_{\mathbb{F}}(Q_1(x) + \cdots + Q_k(x)),$$

where the coefficients $c_{Q_1,\dots,Q_k}$ are complex numbers of magnitude at most B for some $B = O_{k,\dim(\Sigma)}(1)$. We may use this expansion to split f as $f_1 + f_2$, where

(6.1)
$$f_1(x) := \sum_{Q_1 \in \mathrm{Span}(\mathcal{F}_1),\dots,Q_d \in \mathrm{Span}(\mathcal{F}_d)} c_{Q_1,\dots,Q_d,0,\dots,0} e_{\mathbb{F}}(Q_1(x) + \cdots + Q_d(x))$$

and

$$f_2(x) := \sum_{\substack{Q_1 \in \mathrm{Span}(\mathcal{F}_1),\dots,Q_k \in \mathrm{Span}(\mathcal{F}_k) \\ Q_s \neq 0 \text{ for some } s > d}} c_{Q_1,\dots,Q_k} e_{\mathbb{F}}(Q_1(x) + \cdots + Q_k(x)). \tag{6.2}$$

Thus f_2 is the part of f which "genuinely has degree larger than d". We shall show the U^{d+1}-norm of this part is small.

Suppose that polynomials $Q_1 \in \mathrm{Span}(\mathcal{F}_1), \dots, Q_k \in \mathrm{Span}(\mathcal{F}_k)$ are such that Q_s is non-zero and $Q_{s+1}, \dots, Q_{k-1}$ all vanish for some $s > d$. Since $\mathcal{F}$ is F-regular, we have $\mathrm{rank}_{s-1}(Q_s) \geqslant F(\dim(\mathcal{F}))$, and thus

$$\mathrm{rank}_{s-1}(Q_1 + \cdots + Q_k - Q) \geqslant F(\dim(\mathcal{F})) - 1 \tag{6.3}$$

for any $Q \in \mathcal{P}_d(\mathbb{F}^n)$. Applying Theorem 1.8 and the induction hypothesis, we conclude (if F is large enough) that

$$\|e_{\mathbb{F}}(Q_1 + \cdots + Q_k)\|_{U^{k+1}} \leqslant \frac{\delta}{2B|\mathcal{F}_1| \cdots |\mathcal{F}_k|}.$$

Since the Gowers U^{k+1}-norm obeys the triangle inequality (see e.g. [9, Lemma 3.9]), it follows that $\|f_2\|_{U^{k+1}} \leqslant \delta/2$. Recalling that $\|f\|_{U^{k+1}} \geqslant \delta$, another application of the triangle inequality implies that $\|f_1\|_{U^{k+1}} \geqslant \delta/2$. Now by Cauchy-Schwarz we have

$$\|f_1\|_{U^{k+1}}^{2^{k+1}} \leqslant \|f_1\|_2^2 \|f_1\|_\infty^{2^{k+1}-2}.$$

From the bounds on the Fourier coefficients $c_{Q_1,\dots,Q_k}$ we have $\|f_1\|_\infty \ll_{k,\dim(\mathcal{F})} 1$, and therefore

$$\langle f_1, f_1\rangle = \|f_1\|_2^2 \gg_{d,k,\delta,\dim(\mathcal{F})} 1.$$

From (6.1) and the pigeonhole principle it follows that there exist $Q_1 \in \mathcal{F}_1, \dots, Q_d \in \mathcal{F}_d$ such that

$$|\langle f_1, e_{\mathbb{F}}(Q_1 + \cdots + Q_d)\rangle| \geqslant \varepsilon$$

for some $\varepsilon \gg_{d,k,\delta,\dim(\Sigma)} 1$. On the other hand, from (6.3), Theorem 1.8, and (6.2) we have

$$|\langle f_2, e_{\mathbb{F}}(Q_1 + \cdots + Q_d)\rangle| \leqslant \varepsilon/2$$

if F grows sufficiently rapidly. Hence from one further application of the triangle inequality we have

$$|\langle f, e_{\mathbb{F}}(Q_1 + \cdots + Q_d)\rangle| \geqslant \varepsilon/2,$$

and thus $\|f\|_{u^d} \geqslant \varepsilon/2$. Therefore the induction goes through and we have proved Theorem 1.4. □

7. A recurrence result

Proposition 5.1 had a rather lengthy proof. However, the claim is much simpler in the case when the factor $\mathcal{F}$ is trivial. More precisely, we have the following slight generalization of [18, Proposition 4.5].

Lemma 7.1 (Non-zero polynomials do not vanish almost everywhere)**.** *Suppose that $P \in \mathcal{P}_d(\mathbb{F}^n)$ and that $\mathbb{P}_{x\in\mathbb{F}^n}(P(x) = 0) > 1 - 2^{-d}$. Then P is identically zero.*

Remark: This lemma is almost certainly folkloric, but we do not have a precise reference for it.

Proof. We proceed by induction on d, the result being obvious for $d = 1$. For any fixed h we have $\mathbb{P}_{x\in\mathbb{F}^n}(P(x+h) = P(x) = 0) > 1 - 2^{-(d-1)}$. Applying the inductive hypothesis to $P(x+h) - P(x) \in \mathcal{P}_{d-1}(\mathbb{F}^n)$, we see that $P(x+h) - P(x) = 0$ for all x, h. This manifestly implies the result. □

A short consequence of Lemma 7.1 is the following curious recurrence result.

Lemma 7.3 (Multiple polynomial recurrence)**.** *Suppose that $d, k \geqslant 1$ are integers, that $P_1, \dots, P_k \in \mathcal{P}_d(\mathbb{F}^n)$ are polynomials and that $x_0 \in \mathbb{F}^n$. Then*

$$\mathbb{P}_{x\in\mathbb{F}^n}(P_i(x) = P_i(x_0) \text{ for all } i = 1, \dots, k) \geqslant 2^{-(|\mathbb{F}|-1)kd}.$$

Proof. Consider the polynomial

$$Q(x) := \prod_{i=1}^{k} \prod_{\substack{t\in\mathbb{F} \\ t\neq P_i(x_0)}} (P_i(x) - t).$$

This polynomial has degree $(|\mathbb{F}| - 1)kd$, and clearly $Q(x_0) \neq 0$. Applying Lemma 7.1 in the contrapositive, we conclude

$$\mathbb{P}_x(Q(x) \neq 0) \geqslant 2^{-(|\mathbb{F}|-1)kd}$$

and the claim follows. □

Remark: In the case $d < |\mathbb{F}|$, one could also obtain a qualitative version of Lemma 7.3 by combining Lemma 2.4 (applied to the factor generated by $P_1, \dots, P_k$) followed by Lemma 4.1. Of course, the bounds obtained by this approach are far weaker.

8. Representations that respect degree

The results of this section and the next are somewhat technical, and by necessity some of the notation is a little fearsome. First-time readers may wish to skip to the discussion of the counterexample of Theorem 1.5, which is presented in §10.

In previous sections we showed discussed the notion of low-rank polynomials $P \in \mathcal{P}_d(\mathbb{F}^n)$, which can be expressed as $B(Q_1, \dots, Q_k)$ with $Q_i \in \mathcal{P}_{d-1}(\mathbb{F}^n)$. In this section we show how (under a regularity assumption on the factor generated by the Q_i) the function B can be chosen to be a polynomial with controlled degree.

Definition 8.1. *Let $\mathcal{F}$ be a factor of degree $d \geqslant 1$ on a $\mathbb{F}^n$. A* $\mathcal{F}$-monomial *is any product of the form $\prod_{j=1}^{J} Q_j$, where each Q_j belongs to one of the vector spaces* $\mathrm{Span}(\mathcal{F}_{d_j})$ *for some $d_j \in \{1, \dots, d\}$. The* $\mathcal{F}$-degree *of the $\mathcal{F}$-monomial $\prod_{j=1}^{J} Q_j$ is defined to be $\sum_{j=1}^{J} d_j$. If $D \geqslant 0$, we define a* $\mathcal{F}$-polynomial *of $\mathcal{F}$-degree at most D to be any linear combination of $\mathcal{F}$-monomials of $\mathcal{F}$-degree at most D.*

Example 8.2. *Let $\mathbb{F}$ have large characteristic. If $\mathcal{F}$ is the degree 2 factor on $\mathbb{F}^5$ consisting of the four polynomials $X_1X_2 + X_3$, $X_1X_2 + X_4$, $X_2 + X_3$ and $X_1 + X_5$, where $X_1, \dots, X_5$ are the coordinate functions, the polynomial $(X_1X_2 + X_3)(X_1 + X_5)^7 + (X_1 + X_2 + X_3 + X_5)^9$ has $\mathcal{F}$-degree 9, and so does $(X_3 - X_4)^4(X_2 + X_3)$, since $X_3 - X_4 \in \mathrm{Span}(\mathcal{F}_2)$.*

In the above example we saw that the $\mathcal{F}$-degree of a polynomial can exceed the ordinary degree due to dependencies among the polynomials in the factor. The following theorem can be viewed as a converse to this phenomenon.

Theorem 8.3 (Degree and $\mathcal{F}$-degree agree for regular factors)**.** *Let $0 \leqslant d, D < |\mathbb{F}|$. Then there exists a growth function F (depending on d and D) with the following property. Suppose that $P \in \mathcal{P}_D(\mathbb{F}^n)$ is measurable with respect to $\sigma(\mathcal{F})$, where $\mathcal{F}$ is an F-regular factor of degree d on $\mathbb{F}^n$. Then P has $\mathcal{F}$-degree at most D.*

Proof. Let d, D be as above, let F be a rapid growth function to be chosen later, and let $P, \mathcal{F}$ be as above. Since P is measurable with respect to $\sigma(\mathcal{F})$, we have a representation

$$P = B(P_{1,1}, \dots, P_{1,M_1}, \dots, P_{d,1}, \dots, P_{d,M_d})$$

for some function $B : \Sigma \to \mathbb{F}$. As $\mathbb{F}$ is a finite field, we can view B as a polynomial of $\dim(\mathcal{F})$ variables, which has individual degree at most $|\mathbb{F}| - 1$ in each of the variables (note that all higher degrees can be eliminated since $x^{|\mathbb{F}|} = x$). Thus we can write

$$P = \sum_{r \in R} c_r \prod_{i=1}^{d} \prod_{j=1}^{M_i} P_{i,j}^{r_{i,j}}, \tag{8.1}$$

where R is the set of all tuples $r = (r_{i,j})_{1 \leqslant i \leqslant d; 1 \leqslant j \leqslant M_i}$, and the c_r are coefficients in $\mathbb{F}$.

For each tuple $r \in R$, we define the *weight* $|r|$ of r by the formula

$$|r| := \sum_{i=1}^{d} i \sum_{j=1}^{M_i} r_{i,j}.$$

To prove the claim, it suffices to show that $c_r = 0$ for all tuples r with weight larger than D. Suppose for contradiction that this is not the case. Then we can find r with $|r| > D$ such that $c_r \neq 0$; without loss of generality we may assume that $|r|$ is maximal with respect to this property. From (8.1), we thus have

$$P(x) = c_r \prod_{i=1}^{d} \prod_{j=1}^{M_i} P_{i,j}(x)^{r_{i,j}} + \sum_{\substack{s \in R \setminus \{r\} \\ |s| \leqslant |r|}} c_s \prod_{i=1}^{d} \prod_{j=1}^{M_i} P_{i,j}(x)^{s_{i,j}}$$

for all $x \in \mathbb{F}^n$. Since P has degree $D < |r|$, its $|r|^{\text{th}}$ order derivatives vanish. Thus we have

$$0 = c_r \sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} \prod_{i=1}^{d} \prod_{j=1}^{M_i} P_{i,j}(x + \omega \cdot h)^{r_{i,j}} + \sum_{\substack{s \in R \setminus \{r\} \\ |s| \leqslant |r|}} \sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} c_s \prod_{i=1}^{d} \prod_{j=1}^{M_i} P_{i,j}(x + \omega \cdot h)^{s_{i,j}}$$

for all $x \in \mathbb{F}^n$ and $h \in (\mathbb{F}^n)^{|r|}$.

Now if $a = (a_{i,j}(\omega)) \in \Sigma^{\{0,1\}^{|r|}}$ satisfies the parallelelepiped constraints, and if F grows sufficiently rapidly, then we know from Proposition 4.6 that there are $x \in \mathbb{F}^n$ and $h \in (\mathbb{F}^n)^{|r|}$ such that $P_{i,j}(x + \omega \cdot h) = a_{i,j}(\omega)$ for all i, j with $i \in [d]$ and $j \leqslant M_i$ and for all $\omega \in \{0,1\}^{|r|}$. We thus conclude that

$$0 = c_r \sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} \prod_{i=1}^{d} \prod_{j=1}^{M_i} a_{i,j}(\omega)^{r_{i,j}} + \sum_{\substack{s \in R \setminus \{r\} \\ |s| \leqslant |r|}} \sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} c_s \prod_{i=1}^{d} \prod_{j=1}^{M_i} a_{i,j}(\omega)^{s_{i,j}}$$

for all $a \in \Sigma^{\{0,1\}^{|r|}}$ satisfying the parallelepiped constraints. Thus, to obtain the desired contradiction, it will suffice to locate such an a for which

$$\sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} \prod_{i=1}^{d} \prod_{j=1}^{M_i} a_{i,j}(\omega)^{r_{i,j}} \neq 0, \tag{8.2}$$

but such that

$$\sum_{\omega \in \{0,1\}^{|r|}} (-1)^{|\omega|} \prod_{i=1}^{d} \prod_{j=1}^{M_i} a_{i,j}(\omega)^{s_{i,j}} = 0 \tag{8.3}$$

for all $s \in R \setminus \{r\}$ with $|s| \leqslant |r|$.

We can do this explicitly as follows. Let us parametrise $\{0,1\}^{|r|}$ as $\prod_{i=1}^{d} \prod_{j=1}^{M_i} (\{0,1\}^i)^{r_{i,j}}$, thus we write each $\omega \in \{0,1\}^{|r|}$ as $\omega_{i,j,k,t}$, where $1 \leqslant i \leqslant d$, $1 \leqslant j \leqslant M_i$, $1 \leqslant k \leqslant i$ and $1 \leqslant t \leqslant r_{i,j}$. Define $a \in \Sigma^{\{0,1\}^{|r|}}$ by

$$a_{i,j}(\omega) := \sum_{t=1}^{r_{i,j}} \prod_{k=1}^{i} \omega_{i,j,k,t},$$

where we embed $\{0,1\}$ into $\mathbb{F}$ in the obvious way. Since $a_{i,j}(\omega)$ is a linear combination of products of i coordinates of ω, it is easy to see that a satisfies the parallelepiped constraints.

Let us now verify (8.3). For fixed i,j, $a_{i,j}(\omega)$ depends only on the components lying in $(\{0,1\}^i)^{r_{i,j}}$, which are disjoint as i,j vary. We can therefore factorise the left-hand side of (8.3) (with a hopefully obvious notation) as

$$\prod_{i=1}^{d}\prod_{j=1}^{M_i}\Big(\sum_{\eta\in(\{0,1\}^i)^{r_{i,j}}}(-1)^{|\eta|}a_{i,j}(0,\dots,0,\eta,0,\dots,0)^{s_{i,j}}\Big),$$

where the notation is supposed to suggest that η is in the i,j-part of the product $\prod_{i=1}^{d}\prod_{j=1}^{M_i}(\{0,1\}^i)^{r_{i,j}}$. On the other hand, If $|s|\leqslant|r|$ and $s\neq r$, then from the pigeonhole principle there must be some $i\leqslant d$ and some $j\leqslant M_i$ such that $s_{i,j}<r_{i,j}$. Fixing this i,j, it thus suffices to show that

$$\sum_{\eta\in(\{0,1\}^i)^{r_{i,j}}}(-1)^{|\eta|}a_{i,j}(0,\dots,0,\eta,0,\dots,0)^{s_{i,j}}=0.$$

But we observe that $a_{i,j}(\omega)^{s_{i,j}}$ is a linear combination of products of $is_{i,j}$ coordinates of ω, which is strictly less than $ir_{i,j}$, and the claim follows.

Now we verify (8.2). Performing the same factorisation as before, it suffices to show that

$$\sum_{\eta\in(\{0,1\}^i)^{r_{i,j}}}(-1)^{|\eta|}a_{i,j}(0,\dots,0,\eta,0,\dots,0)^{r_{i,j}}\neq 0 \tag{8.4}$$

for each i,j. But $a_{i,j}(0,\dots,0,\eta,0,\dots,0)^{r_{i,j}}$ is equal to $r_{i,j}!\prod_{k=1}^{i}\prod_{t=1}^{r_{i,j}}\eta_{k,t}$ (viewed of course as an element of $\mathbb{F}$), plus several other monomials, none of which involve all of the $\eta_{k,t}$. From this we see that the left-hand side of (8.4) is simply $(-1)^{ir_{i,j}}r_{i,j}!$. Since $r_{i,j}<|\mathbb{F}|$, this expression is non-zero in $\mathbb{F}$, as desired. □

Combining this theorem with Lemma 2.4 we immediately obtain the following corollary.

Corollary 8.4 (Minimal-degree representation of polynomials)**.** *Let $1\leqslant d,D<|\mathbb{F}|$, and let F be a growth function. Then whenever $P\in\mathcal{P}_D(\mathbb{F}^n)$ is measurable with respect to a factor $\mathcal{F}$ of degree d on $\mathbb{F}^n$, there exists an F-regular extension $\mathcal{F}'$ of $\mathcal{F}$ of degree d with $\dim(\mathcal{F}')\ll_{d,D,\dim(\mathcal{F})}1$ such that P has $\mathcal{F}'$-degree at most D.*

9. A nullstellensatz

In this section we establish a kind of finite field analogue of Hilbert's Nullstellensatz. These results are not needed elsewhere in the paper, but are illustrative applications of the previous machinery, and may be of some independent interest.

Proposition 9.1 (Nullstellensatz)**.** *Let $k\geqslant 0$ and $0\leqslant d<|\mathbb{F}|$, and let $P_1,\dots,P_k\in\mathcal{P}_d(\mathbb{F}^n)$. Let $Q\in\mathcal{P}_d(\mathbb{F}^n)$ be such that Q vanishes whenever*

$P_1, \dots, P_k$ all vanish. Then there exist polynomials $R_1, \dots, R_k$ of degree $O_{d,k}(1)$ such that

$$Q(x) = P_1(x)R_1(x) + \cdots + P_k(x)R_k(x)$$

for all $x \in \mathbb{F}^n$.

Proof. Let $\mathcal{F}$ be the degree d factor defined by the polynomials $P_1, \dots, P_k, Q$. Let F be a growth function to be chosen later. By Lemma 2.4, we can extend $\mathcal{F}$ to an F-regular factor $\mathcal{F}'$ of degree d and dimension $O_{d,k,F}(1)$. If F is sufficiently rapid, then by Lemma 4.1 we see that the configuration map $\Phi' : \mathbb{F}^n \to \Sigma'$ corresponding to $\mathcal{F}'$ is surjective. Since $P_1, \dots, P_k, Q$ are measurable with respect to $\sigma(\mathcal{F}')$, we can write $P_i = p_i \circ \Phi'$ and $Q = q \circ \Phi'$ for some $p_i, q : \Sigma' \to \mathbb{F}$. Our assumption together with the surjectivity of Φ' implies that if $z \in \Sigma'$ is such that $p_i(z) = 0$ for $i = 1, \dots, k$ then $q(z) = 0$. By working on each point z separately, one can therefore find functions $r_1, \dots, r_k : \Sigma' \to \mathbb{F}$ such that

$$q(z) = p_1(z)r_1(z) + \cdots + p_k(z)r_k(z)$$

for all $z \in \Sigma'$. Composing with Φ' we conclude that

$$Q(x) = P_1(x)R_1(x) + \cdots + P_k(x)R_k(x)$$

for all $x \in \mathbb{F}^n$, where $R_i := r_i \circ \Phi'$. As Σ' has dimension $O_{d,k,F}(1)$, one can view $r_1, \dots, r_k$ as polynomials of degree $O_{d,k,F}(1)$, and so $R_1, \dots, R_k$ are also polynomials of degree $O_{d,k,F}(1)$. The claim follows. □

In the above result the polynomials R_i had bounded degree. However, if the polynomials $P_1, \dots, P_k$ arose from a sufficiently regular factor, one can get the sharp degree bound for R_i, namely $\deg(R_i) = \deg(Q) - \deg(P_i)$.

Proposition 9.2 (Exact nullstellensatz)**.** *Let $D, d, k \geqslant 0$. Then there exists a growth function F (depending on D, d, k) with the following property: given any F-regular factor $\mathcal{F}$ of degree d and dimension at most D on $\mathbb{F}^n$, and given any $Q \in \mathcal{P}_k(\mathbb{F}^n)$ which vanishes whenever the polynomials $P_{i,j}$ defining $\mathcal{F}$ all vanish, there exist polynomials $R_{i,j} \in \mathcal{P}_{k-i}(\mathbb{F}^n)$ for all $i \leqslant \min(d,k)$ and $j \leqslant M_i$ such that*

$$Q(x) = \sum_{i=1}^{\min(d,k)} R_{i,j}(x)P_{i,j}(x)$$

for all $x \in \mathbb{F}^n$.

Before embarking on the proof, we give a technical generalisation of the regularity lemma, Lemma 2.4. Let us say that an extension $\mathcal{F}'$ of a factor $\mathcal{F}$ of degree d is *non-disruptive* if we have $\mathcal{F}_i \subseteq \mathcal{F}'_i$ for all $i = 1, \dots, d$. Clearly if $\mathcal{F}'$ is a non-disruptive extension of $\mathcal{F}$ and $\mathcal{F}'$ is F-regular, then $\mathcal{F}$ must also be F-regular. Our next lemma can be regarded as a kind of converse to this fact.

Lemma 9.3 (Relative regularity lemma)**.** *Let $d, D \geqslant 1$ and let F be a growth function. Then there exists a growth function $\tilde{F}$ such that whenever $\mathcal{F}$ is a $\tilde{F}$-regular factor of degree d on $\mathbb{F}^n$, and $\mathcal{F}'$ is an extension of $\mathcal{F}$ of dimension at most D, there exists an F-regular extension $\mathcal{F}''$ of $\mathcal{F}'$ with the dimension bound*

$$\dim(\mathcal{F}'') \ll_{F,d,D} 1 \tag{9.1}$$

such that $\mathcal{F}''$ is a non-disruptive extension of $\mathcal{F}$.

Proof. Fix d, F, and let $\tilde{F}$ be a sufficiently rapid growth function to be chosen later. First observe that as the polynomials in $\mathcal{F}$ are $\mathcal{F}'$-measurable, we have the crude bound $\dim(\mathcal{F}) \ll_D 1$, and so we may allow our constants to depend on $\dim(\mathcal{F})$ also.

By replacing $\mathcal{F}'_i$ with $\mathcal{F}'_i \cup \mathcal{F}_i$ for $1 \leqslant i \leqslant d$ if necessary (and increasing D accordingly) we may assume that $\mathcal{F}'$ is a non-disruptive extension of $\mathcal{F}$. We now keep $\mathcal{F}$ fixed and induct on the dimension vector $(\dim(\mathcal{F}'_1), \dots, \dim(\mathcal{F}'_d))$ of $\mathcal{F}'$ in exactly the same way as in Lemma 2.4 in order to obtain an F-regular extension $\mathcal{F}''$ of $\mathcal{F}'$ obeying (9.1). The key point is that the low-rank polynomials Q_i which arise in the proof of Lemma 2.4 can never arise from $\mathcal{F}_i$ if $\tilde{F}$ is chosen sufficiently rapid (thanks to (9.1)). Because of this, we can easily arrange that the extension $\mathcal{F}''$ appearing in the proof of Lemma 2.4 continues to be a non-disruptive extension of $\mathcal{F}$, and the claim easily follows. □

Proof of Proposition 9.2. Fix $D, d, k \geqslant 0$. By adding dummy polynomials to $\mathcal{F}$ and enlarging d if necessary we may assume that $d \geqslant k$. Let F_1 be a growth function depending on D, d, k to be chosen later, and let F be an even more rapid growth function depending on D, d, k, F_1 and also to be chosen later.

Let $\mathcal{F}, Q$ be as in the statement of the proposition. Let $\mathcal{F}' = (P_{i,j})_{i\in[d], j\leqslant M'_i}$ be the factor of degree d formed by adjoining Q to $\mathcal{F}$. Applying Lemma 9.3, we see (if F is sufficiently rapid depending on D, d, k, F_1) that we can find an F_1-regular extension $\mathcal{F}'' = (P_{i,j})_{i\in[d], j\leqslant M''_i}$ of $\mathcal{F}$ of degree $\max(d, k)$ which is a non-disruptive extension of $\mathcal{F}$. Applying Theorem 8.3, we conclude (if F_1 is sufficiently rapid depending on D, d, k) that Q has $\mathcal{F}''$-degree at most k. Using the identity $x^{|\mathbb{F}|} = x$ to eliminate all exponents greater than or equal to $|\mathbb{F}|$, we have a representation $Q(x) = q(\Phi''(x))$ for all $x \in \mathbb{F}^n$, where $q : \Sigma'' \to \mathbb{F}$ is a polynomial which takes the form

$$q(t) := \sum_{s\in S_k} c_s \prod_{i=1}^{d} \prod_{j=1}^{M''_i} t_{i,j}^{s_{i,j}} \tag{9.2}$$

($c_s \in \mathbb{F}$ for all $s \in S_k$), and S_k is the collection of all tuples $(s_{i,j})_{1 \leqslant i \leqslant d; 1 \leqslant j \leqslant M_i''}$ of non-negative integers $0 \leqslant s_{i,j} < |\mathbb{F}|$ obeying the weight condition

$$\sum_{i=1}^{d} \sum_{j=1}^{M_i''} i s_{i,j} \leqslant k.$$

By hypothesis, $Q(x)$ vanishes whenever all the $P_{i,j}(x)$ vanish for $i = 1, \ldots, d$ and $j \leqslant M_i$. On the other hand, by Lemma 4.1 we see (if F_1 is sufficiently rapid) that $\Phi'' : \mathbb{F}^n \to \Sigma''$ is surjective. We conclude that q vanishes on the coordinate subspace

$$W := \{t \in \Sigma'' : t_{i,j} = 0 \text{ for all } i = 1, \ldots, d \text{ and } j \leqslant M_i'\}.$$

Restricting q to W and then equating coefficients (recalling from the Lagrange interpolation formula that the coefficients are uniquely determined as long as all exponents are less than $|\mathbb{F}|$) we conclude that c_s vanishes for each $s \in S$ such that $s_{i,j} = 0$ for all i, j with $i \leqslant d$ and $j \leqslant M_i$. From this, we can easily obtain a representation of the form

$$q(t) = \sum_{i=1}^{d} \sum_{j=1}^{M_i} t_{i,j} r_{i,j}(t),$$

where each $r_{i,j}$ has weighted degree at most $k - i$ in the sense that it can be expanded into monomials as in (9.2) but using only exponents from S_{k-i} rather than all of S_k. In particular $r_{i,j}$ must vanish for $i > k$. Substituting $t = \Phi''(x)$ we obtain the claim. □

10. The counterexample

In this section we analyse the counterexample to the inverse conjecture for the Gowers norms in characteristic two by proving Theorem 1.5. Recall what is claimed in that theorem: the elementary symmetric quartic

$$S_4(x) = \sum_{1 \leqslant i_1 < i_2 < i_3 < i_4 \leqslant n} x_{i_1} x_{i_2} x_{i_3} x_{i_4}$$

is such that $f(x) = (-1)^{S_4(x)}$ has large U^4-norm on $\mathbb{F}_2^n$, but this function does not correlate well with any cubic phase.

We begin by establishing that the U^4-norm of this function is large. Define the symmetric bilinear form $B : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ by

$$B(a, b) := \sum_{1 \leqslant i,j \leqslant n : i \neq j} a_i b_j \tag{10.1}$$

for $a = (a_1, \dots, a_n), b = (b_1, \dots, b_n)$ in $\mathbb{F}_2^n$. One readily verifies[3] the identity

$$\begin{aligned} D_a D_b D_c D_d S_4(x) &= \sum_{\substack{1 \leqslant i,j,k,l \leqslant n \\ i,j,k,l \text{ distinct}}} a_i b_j c_k d_l \\ &= B(a,b)B(c,d) + B(a,c)B(b,d) + B(a,d)B(b,c), \end{aligned} \tag{10.2}$$

and so

$$\|f\|_{U^4}^4 = \mathbb{E}_{a,b,c,d \in \mathbb{F}_2^n} (-1)^{B(a,b)B(c,d)+B(a,c)B(b,d)+B(a,d)B(b,c)}. \tag{10.3}$$

To compute this quantity, we need to look at the distribution of the sextuplet

$$B_6(a,b,c,d) := (B(a,b), B(a,c), B(a,d), B(b,c), B(b,d), B(c,d)) \in \mathbb{F}_2^6 \tag{10.4}$$

as a, b, c, d vary in $\mathbb{F}_2^n$. This distribution can be controlled by standard Gauss sum estimates such as the following (cf. also Lemma 1.7).

Lemma 10.1 (Gauss sum estimate)**.** *For any $\xi_{ab}, \xi_{ac}, \xi_{ad}, \xi_{bc}, \xi_{bd}, \xi_{cd} \in \mathbb{F}_2$, not all zero, we have*

$$\mathbb{E}_{a,b,c,d \in \mathbb{F}_2^n} (-1)^{\xi_{ab}B(a,b)+\xi_{ac}B(a,c)+\xi_{ad}B(a,d)+\xi_{bc}B(b,c)+\xi_{bd}B(b,d)+\xi_{cd}B(c,d)} = O(2^{-n/2}).$$

Proof. By symmetry we may assume $\xi_{ab} = 1$. It suffices to show that

$$\mathbb{E}_{a,b \in \mathbb{F}_2^n} (-1)^{B(a,b)+\xi_{ac}B(a,c)+\xi_{ad}B(a,d)+\xi_{bc}B(b,c)+\xi_{bd}B(b,d)+\xi_{cd}B(c,d)} = O(2^{-n/2})$$

uniformly in $c, d \in \mathbb{F}_2^n$. But if we fix c, d, we can write the left-hand side as

$$\mathbb{E}_{a,b \in \mathbb{F}_2^n} (-1)^{B(a,b)+L(a)+L'(b)}$$

for some $L, L' \in \mathcal{P}_1(\mathbb{F}_2^n)$. Applying Cauchy-Schwarz to eliminate the $(-1)^{L'(b)}$ factor, we can estimate this quantity in absolute value by

$$|\mathbb{E}_{a,a',b \in \mathbb{F}_2^n} (-1)^{B(a,b)-B(a',b)+L(a)-L(a')}|^{1/2};$$

writing $c := a - a'$ this becomes

$$|\mathbb{E}_{c,b \in \mathbb{F}_2^n} (-1)^{B(c,b)+L(c)}|^{1/2}.$$

Performing the c average using Fourier analysis and using the triangle inequality, we can bound this by

$$|\mathbb{P}_{c \in \mathbb{F}_2^n}(B(c,b) = 0 \text{ for all } b \in \mathbb{F}_2^n)|^{1/2}.$$

But B has rank $n - O(1)$, and so

$$\mathbb{P}_{c \in \mathbb{F}_2^n}(B(c,b) = 0 \text{ for all } b \in \mathbb{F}_2^n) = O(2^{-n}).$$

The claim follows. □

[3]For a generalisation of this identity, see Lemma 11.2 below.

From this lemma and Fourier analysis on $\mathbb{F}_2^6$ (as in the proof of Lemma 4.1) we see that B_6 is equidistributed in the sense that

$$\mathbb{P}_{a,b,c,d\in\mathbb{F}_2^n}(B_6(a,b,c,d)=q)=2^{-6}+O(2^{-n}) \text{ for all } q\in\mathbb{F}_2^6.$$

It follows that (10.3) can be rewritten as

$$\mathbb{E}_{q_{ab},q_{ac},q_{ad},q_{bc},q_{bd},q_{cd}\in\mathbb{F}_2}(-1)^{q_{ab}q_{cd}+q_{ac}q_{bd}+q_{ad}q_{bc}}+O(2^{-n}).$$

But we can factorise the expectation and rewrite this expression as

$$(\mathbb{E}_{q,q'\in\mathbb{F}_2}(-1)^{qq'})^3+O(2^{-n/2}).$$

Since $\mathbb{E}_{q,q'\in\mathbb{F}_2}(-1)^{qq'}=\frac{1}{2}$, it follows that $\|f\|_{U^4}^4=\frac{1}{8}+O(2^{-n})$ as asserted in (1.4) of Theorem 1.5.

Now we turn to (1.5), which asserts that f does not have substantial correlation with a cubic phase. Let us remind the reader once more that a better bound is contained in the independent work of Lovett, Meshulam and Samorodnitsky [15]. Our bound is all but contained in Alon and Beigel [2, Theorem 7], although we recall that argument here for the convenience of the reader.

If $x=(x_1,\dots,x_n)\in\mathbb{F}_2^n$, let $|x|$ denote the number of indices $i\in[n]$ for which $x_i=1$. It is clear that $S_d(x)=\binom{|x|}{d} \pmod 2$. Recalling Lucas' theorem on binomial coefficients $\pmod p$, which states that

$$\binom{a}{b}\equiv\binom{a_0}{b_0}\cdots\binom{a_k}{b_k} \pmod p \tag{10.5}$$

whenever $a=a_0+a_1p+a_2p^2+\cdots+a_kp^k$ and $b=b_0+b_1p+b_2p^2+\cdots+b_kp^k$ with $0\leqslant a_i,b_i<p$, we see that

$$\begin{aligned}
&S_0(x)=0\\
&S_1(x)=1 \text{ iff } |x|\equiv 1 \pmod 2\\
&S_2(x)=1 \text{ iff } |x|\equiv 2,3 \pmod 4\\
&S_3(x)=1 \text{ iff } |x|\equiv 3 \pmod 4 \text{ and}\\
&S_4(x)=1 \text{ iff } |x|\equiv 4,5,6,7 \pmod 8.
\end{aligned}$$

On the other hand we have, by a technique once known[4] as "multisection of series",

$$\begin{aligned}
\mathbb{P}_{x\in\mathbb{F}_2^n}(|x|\equiv a \pmod 8) &= 2^{-n}\sum_{j\equiv a \pmod 8}\binom{n}{j}\\
&=\frac{1}{8}\sum_{r=0}^{7}e^{-2\pi ira/8}\left(\frac{1+e^{2\pi ir/8}}{2}\right)^n\\
&=\frac{1}{8}+O(2^{-\Omega(n)}).
\end{aligned}$$

[4]One can also interpret this computation as exhibiting (by the usual Fourier-analytic method) the exponential mixing rate of a simple random walk on $\mathbb{Z}/8\mathbb{Z}$.

From these facts and some computation we easily conclude that

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0} = O(2^{-\Omega(n)})$$

for all coefficients $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$. Clearly this immediately implies that

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_0} = O(2^{-\Omega(n)}) \tag{10.6}$$

whenever $Q_0 \in \mathcal{P}_0(\mathbb{F}_2^n)$ and $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$.

Now suppose instead that $Q_1 \in \mathcal{P}_1(\mathbb{F}_2^n)$ and $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$, and consider the average

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_1}. \tag{10.7}$$

Then we can write

$$Q_1(x) = \sum_{i\in E} x_i + Q_0(x)$$

for some $Q_0 \in \mathcal{P}_0(\mathbb{F}_2^n)$ and some set $E \subset \{1,\dots,n\}$. We can thus find a set $I \subseteq \{1,\dots,n\}$ of size $m := \lfloor \frac{n}{2} \rfloor$ which either lies in E, or is disjoint from E. By permuting the coefficients we can write $I = \{1,\dots,m\}$. Then by freezing the coefficients $y := (x_{m+1},\dots,x_n) \in \mathbb{F}_2^{n-m}$, we see that we can write (10.7) as an average of expressions of the form

$$\mathbb{E}_{x\in\mathbb{F}_2^m}(-1)^{S_4(x)+c_{3,y}S_3(x)+c_{2,y}S_2(x)+c_{1,y}S_1(x)+c_{0,y}S_0-Q_{0,y}}$$

for some $c_{0,y},\dots,c_{3,y} \in \mathbb{F}_2$ and $Q_{0,y} \in \mathcal{P}_1(\mathbb{F}_2^m)$. Applying (10.6) and the triangle inequality we thus conclude that

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_1(x)}) = O(2^{-\Omega(n)}). \tag{10.8}$$

Now suppose instead that $Q_2 \in \mathcal{P}_2(\mathbb{F}_2^n)$ and $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$, and consider the average

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_2(x)}.$$

Then we can write

$$Q_2(x) = \sum_{\{i,j\}\in E(\Gamma)} x_ix_j + Q_1(x)$$

for some $Q_1 \in \mathcal{P}_1(\mathbb{F}_2^n)$ and some graph Γ on vertex set $[n]$. By Ramsey's theorem (see e.g. [10, Section 4.2]), we can find a set $I \subseteq [n]$ of size $m = \Omega(\log n)$ such that the complete graph on vertex set I either lies completely inside E, or is disjoint from E. We can then repeat the above freezing argument (using (10.8) instead of (10.6)) and conclude that

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_2(x)} = O(2^{-\Omega(m)}) = O(n^{-\Omega(1)}).$$

Finally, suppose $Q_3 \in \mathcal{P}_1(\mathbb{F}_2^n)$ and $c_0, c_1, c_2, c_3 \in \mathbb{F}_2$, and consider the average

$$\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_3(x)}.$$

Then we can write

$$Q_3(x) = \sum_{\{i,j,k\}\in E(\Gamma)} x_ix_jx_k + Q_2(x)$$

for some $Q_2 \in \mathcal{P}_2(\mathbb{F}_2^n)$ and some 3-uniform hypergraph Γ on vertex set $[n]$. Applying the bounds of Erdős and Rado for the hypergraph Ramsey theorem (see e.g. [10, Section 4.7]) we can find a set $I \subset [n]$ of size $m = \Omega(\log\log n)$ such that the complete 3-uniform hypergraph on I either lies completely inside E or is disjoint from E. Using the freezing argument one last time, we obtain

(10.9)
$$\begin{aligned}\mathbb{E}_{x\in\mathbb{F}_2^n}(-1)^{S_4(x)+c_3S_3(x)+c_2S_2(x)+c_1S_1(x)+c_0S_0-Q_3(x)} &= O(m^{-\Omega(1)})\\ &= O((\log\log n)^{-\Omega(1)}).\end{aligned}$$

This is a bound of the form claimed in (1.5) of Theorem 1.5, except there is an extra logarithm. To remove it, we run the two Ramsey-theoretic arguments in parallel, by using the following variant of the Erdős-Rado bound.

Lemma 10.2 (Simultaneous Ramsey theorem)**.** *Let $E_2 \subseteq \binom{[n]}{2}$ and $E_3 \subseteq \binom{[n]}{3}$ be a graph and 3-uniform hypergraph respectively. Then there exists a set $I \subset [n]$ of size $m = \Omega(\log\log n)$ such that for each $j = 2, 3$, the set $\binom{I}{j}$ either lies completely inside E_j or is disjoint from E_j.*

Proof. We generate some vertices $x_1, \dots, x_l$ by the following algorithm:

Step 0: Initialise $l = 0$ and $J := [n]$.

Step 1: By the pigeonhole principle, there exists $J' \subseteq J$ with $|J'| \gg 2^{-O(l^2)}|J|$ such that for any $i, j \in [l]$ and $x \in J'$, the truth value of the statements $\{x_i, x\} \in E_2$ or $\{x_i, x_j, x\} \in E_3$ are independent of x. Fix this J'.

Step 2: Set $x_{l+1} := \min(J')$, replace J by $J'\backslash\{x_{l+1}\}$, and increment l to $l+1$. If J' is non-empty then return to Step 1; otherwise STOP.

One easily verifies that this algorithm terminates in $k = \Omega(\log^{1/3} n)$ steps to obtain a sequence $1 \leqslant x_1 \leqslant \cdots \leqslant x_l \leqslant n$ with the property that for any $1 \leqslant i < j \leqslant l$, the truth value of $\{x_i, x_j\} \in E_2$ is independent of j, and for any $1 \leqslant i < j < k \leqslant l$, the truth value of $\{x_i, x_j, x_k\} \in E_3$ is independent of k. By an appeal to Ramsey's theorem for graphs one can then find a set $I \subset \{x_1, \dots, x_k\}$ with $|I| \gg \log k \gg \log\log n$ with the desired properties. □

Note that by applying Ramsey's theorem for graphs and 3-uniform hypergraphs sequentially, one would only get $m = \Omega(\log\log\log n)$ here. The reader can easily verify that the logarithmic saving in this lemma propagates through the previous arguments to improve (10.9) to (1.5).

11. General degrees and characteristics

It is natural to wonder for which $\mathbb{F}$ and d the symmetric polynomials S_d on $\mathbb{F}^n$ provide counterexamples to Conjecture 1.3, the inverse conjecture for the U^d-norm. We do not have a complete answer to this question, but we give

some partial results in this direction here. For a more in-depth treatment of these issues, we refer the reader to the recent preprint [15].

We begin with a general result that shows that $\|e_{\mathbb{F}}(S_d)\|_{U^d}$ is large whenever $d > |\mathbb{F}|$. This result (and in fact a generalisation of it which establishes the largeness of $\|e_{\mathbb{F}}(S_d)\|_{U^{d-p+2}}$ for $d \geqslant 2p$, where $p = |\mathbb{F}|$) was shown to us by the authors of [15] before we wrote this section. The following argument is a slight variant of theirs which, we believe, is worth having in the literature.

Theorem 11.1 (Lower bound on Gowers norm)**.** *Let $\mathbb{F}$ be a finite field, let $n \geqslant 1$, and let $d > |\mathbb{F}|$. Let S_d be the symmetric polynomial on $\mathbb{F}^n$, and let $f := e_{\mathbb{F}}(S_d)$. Then $\|f\|_{U^d} \gg_{\mathbb{F}} 1$.*

Proof. For this, we must find some analogue of the computations earlier in the section and, in particular, the identity (10.2). For this we need some more notation. Let Π_n denote the collection of all partitions $\pi = \{C_1, \dots, C_m\}$ of $[n]$ into disjoint sets $[n] = C_1 \cup \dots \cup C_m$. For any partition $\pi = \{C_1, \dots, C_m\} \in \Pi_n$, we associate the multilinear form $R_\pi : \mathbb{F}^n \times \dots \times \mathbb{F}^n \to \mathbb{F}$ by

$$R_\pi(h^{(1)}, \dots, h^{(d)}) := \prod_{k=1}^{m} \sum_{j=1}^{n} \prod_{i \in C_k} h_j^{(i)}.$$

Thus for example if π is the partition of $[3]$ into $\{1,2\}$ and $\{3\}$ then we have

$$R_\pi(h^{(1)}, h^{(2)}, h^{(3)}) = (h_1^{(1)} h_1^{(2)} + \dots + h_n^{(1)} h_n^{(2)})(h_1^{(3)} + \dots + h_n^{(3)}).$$

We define the *Möbius function* $\mu(\pi)$ of μ at π by the formula

$$\mu(\pi) := \prod_k (-1)^{|C_k|}(|C_k| - 1)!. \tag{11.1}$$

We place a partial ordering on partitions π by declaring $\pi' \preceq \pi$ if every set in π' is contained in some set in π. This has a minimal element $\pi_{\min} := \{\{1\}, \dots, \{n\}\}$. The Möbius function can be shown[5] to obey the Möbius inversion identities $\mu(\pi_{\min}) = 1$ and $\sum_{\pi' \preceq \pi} \mu(\pi') = 0$ if $\pi \neq \pi_{\min}$.

As a consequence we obtain the following variant of (10.2), which follows from [15, Proposition 2.7].

Lemma 11.2 (Derivative of symmetric function)**.** *For any $d \geqslant 1$ and $h^{(1)}, \dots, h^{(d)}, x \in \mathbb{F}^n$, we have*

$$D_{h^{(1)}} \dots D_{h^{(d)}} S_d(x) = \sum_\pi \mu(\pi) R_\pi(h^{(1)}, \dots, h^{(d)}). \tag{11.2}$$

Proof. Each R_π may be expanded as a sum

$$R_\pi(h^{(1)}, \dots, h^{(d)}) = \sum_{\pi \preceq \tau(i_1, \dots, i_n)} h_{i_1}^{(1)} \dots h_{i_n}^{(n)}, \tag{11.3}$$

[5] See for instance the series of exercises [4, p. 103], or [7, Lemma 4.1].

where $\tau(i_1,\dots,i_n)$ is the partition on $[n]$ induced by the indices $i_1,\dots,i_n$, two elements s,t being placed in the same element of this partition if and only if $i_s = i_t$.

On the other hand, from proof of Theorem 1.5 we have

$$\begin{aligned} D_{h^{(1)},\dots,h^{(d)}} S_d(x) &= \sum_{\substack{1\leqslant i_1,\dots,i_d\leqslant n \\ i_1,\dots,i_d \text{ distinct}}} h^{(1)}_{i_1}\cdots h^{(d)}_{i_d} \\ &= \sum_{\tau(i_1,\dots,i_n)=\pi_{\min}} h^{(1)}_{i_1}\cdots h^{(n)}_{i_n}. \end{aligned} \tag{11.4}$$

The claim now follows from the Möbius inversion formula. □

To apply the identity (11.2), we let $V \subseteq \mathbb{F}^n$ be the variety

$$V := \{x \in \mathbb{F}_2^n : S_1(x) = S_2(x) = \cdots = S_p(x) = 0\},$$

where $p = |\mathbb{F}|$ (later on we will specialize to the case $p = 2$). We claim the identity

$$\Delta_{h^{(1)}}\cdots\Delta_{h^{(d)}}(f1_V)(x) = \Delta_{h^{(1)}}\cdots\Delta_{h^{(d)}}(1_V)(x) \tag{11.5}$$

for $x, h^{(1)},\dots,h^{(d)} \in \mathbb{F}^n$. To prove (11.5), it suffices to show that

$$D_{h^{(1)}}\cdots D_{h^{(d)}} S_d(x) = 0$$

whenever $x, h^{(1)},\dots,h^{(d)} \in \mathbb{F}^n$ are such that the cube $\{x + \omega_1 h^{(1)} + \cdots + \omega_d h^{(d)} : \omega_1,\dots,\omega_d \in \{0,1\}\}$ lies in V. But if $x, h^{(1)},\dots,h^{(d)}$ are such elements then, by definition of V and differentiation, we have

$$D_{h^{(i_1)}}\cdots D_{h^{(i_j)}} S_j(x) = 0 \tag{11.6}$$

for all $j \in \{1,\dots,p\}$ and distinct $i_1,\dots,i_j \in \{1,\dots,d\}$. Note from (11.1) that the Möbius function $\mu(\pi_{\text{triv},j})$ is invertible in $\mathbb{F}$ for all $1 \leqslant j \leqslant p$, where $\pi_{\text{triv},j}$ is the trivial partition $\{\{1,\dots,j\}\}$ of $[j]$. By expanding the left-hand side of (11.6) using the inversion formula (11.2), we conclude recursively that

$$R_{\pi_{\text{triv},j}}(h^{(i_1)},\dots,h^{(i_j)}) = 0$$

for all $1 \leqslant j \leqslant p$ and distinct $i_1,\dots,i_j \in \{1,\dots,d\}$. This implies that

$$R_\pi(h^{(1)},\dots,h^{(d)}) = 0$$

whenever all sets in π have cardinality at most p. On the other hand, if any set in π has cardinality greater than p, we see from (11.1) that $\mu(\pi)$ vanishes in $\mathbb{F}$. The claim (11.5) now follows from one last application of (11.2).

Using (11.5) and Definition 1.1, we conclude that

$$\|f1_V\|_{U^d} = \|1_V\|_{U^d}.$$

But by monotonicity of Gowers norms (see e.g. [19, Chapter 11]) we have

$$\|1_V\|_{U^d} \geqslant \|1_V\|_{U^1} = |V|/|\mathbb{F}^n|.$$

By applying Lemma 7.3 we have $|V|/|\mathbb{F}^n| \gg_{\mathbb{F}} 1$, and so

$$\|f1_V\|_{U^d} \gg_{\mathbb{F}} 1.$$

On the other hand, we have the Fourier expansion

$$1_V = \mathbb{E}_{\xi \in \mathbb{F}^p} e_{\mathbb{F}}(\xi_1 S_1 + \cdots + \xi_p S_p).$$

Using the triangle inequality for Gowers norms (see e.g. [9, Lemma 3.9] or [19, Chapter 11]) we conclude that

$$\|f e_{\mathbb{F}}(\xi_1 S_1 + \cdots + \xi_p S_p)\|_{U^d} \gg_{\mathbb{F}} 1$$

for some $\xi_1, \ldots, \xi_p \in \mathbb{F}$. Theorem 11.1 now follows from (1.1) and the hypothesis that $d > p$. □

As a consequence of the above theorem, we can completely characterise the behaviour of $(-1)^{S_d}$ in the characteristic 2 case.

Theorem 11.3 (Gowers norm behaviour of S_d over $\mathbb{F}_2$)**.** *Let $n \geqslant 1$ and $d \geqslant 1$ be integers, let $\mathbb{F} = \mathbb{F}_2$, and let $f := (-1)^{S_d}$ where S_d is the d^{th} elementary symmetric function on $\mathbb{F}_2^n$.*

- *If $d = 1, 2$, then $\|f\|_{U^d}, \|f\|_{u^d} = o(1)$.*
- *If d is not a power of 2, then $\operatorname{rank}_{d-1}(S_d) \leqslant 2$ and*
$$\|f\|_{U^d} \geqslant \|f\|_{u^d} \geqslant \frac{1}{4}.$$
- *If d is a power of 2 which is at least 4, then*
$$\|f\|_{U^d} \gg 1 \text{ and } \|f\|_{u^d} = o_d(1),$$
where $o_d(1)$ goes to zero as $n \to \infty$ for fixed d. (In particular, Conjecture 1.3 fails for the U^d-norm on $\mathbb{F}_2^n$ for these values of d.)

Proof. The cases $d = 1, 2$ can be computed by hand (using Lemma 1.7 for the $d = 2$ case). If d is not a power of 2, then from Lucas' theorem (10.5) we can express S_d as a product $S_{d_1} S_{d_2}$ for some d_1, d_2 with $0 < d_1, d_2 < d$ and $d = d_1 + d_2$, which gives the desired bound on $\operatorname{rank}_{d-1}(S_d)$. By Fourier analysis in $\mathbb{F}_2^{k+1}$ we may therefore write

$$(-1)^{S_d} = \frac{1}{4}\left(1 + (-1)^{S_{d_1}} + (-1)^{S_{d_2}} + (-1)^{S_{d_1}+S_{d_2}}\right).$$

Thus $(-1)^{S_d}$ must have an inner product of at least $\frac{1}{4}$ with at least one polynomial phase of degree strictly less than d, which gives the lower bound on $\|f\|_{u^d}$ in this case. The lower bound on $\|f\|_{U^d}$ then follows from (1.2).

When d is a power of 2, one verifies (as in the proof of Theorem 1.5) that $S_d(x) = 1$ precisely when x is equal to $d, \ldots, 2d-1 \pmod{2d}$, whereas $S_{d'}$ for $d' < d$ is periodic with period dividing d. Using multisection of series as before, we can conclude an analogue of (10.6) for S_d instead of S_4, and by repeating the Ramsey arguments one obtains the desired bound $\|f\|_{u^d} = o_d(1)$. Finally, the lower bound on $\|f\|_{U^d}$ follows from Theorem 11.1. This establishes all the claims of the theorem. □

Remark: When $\mathbb{F} = \mathbb{F}_2$ and d is a power of two, the above theorem shows that $(-1)^{S_d}$ does not correlate strongly with any polynomial phase in $\mathbb{F}_2^n$ of degree $d-1$ or less. However, the argument we used to prove this showed that S_d was still *locally polynomial* of degree $d-1$ on the subvariety $V := \{x \in \mathbb{F}_2^n : S_1(x) = S_2(x) = 0\}$, in the sense of [12]. This raises the possibility that Conjecture 1.3 may be salvaged by working with *locally* polynomial phases instead of global ones; in fact this formulation of the conjecture was already implicit in [12, Section 13].

12. Acknowledgements

The authors are indebted to Andrej Bogdanov, Tali Kaufman, and Emanuele Viola for suggesting this problem, and for many useful discussions. The authors also thank Alex Samorodnitsky for drawing attention to the recent preprint [15], and to Peter Sarnak for suggestions.

References

1. N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), no. 1–2, 7–29.
2. N. Alon and R. Beigel, *Lower bounds for approximations by low degree polynomials over Z_m*, Proc. of the 16th Annual IEEE Conference on Computational Complexity (CCC), IEEE, 2001, pp. 184–187.
3. N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron, *Testing low-degree polynomials over GF(2)*, RANDOM-APPROX (2003), 188–199, See also: *Testing Reed-Muller codes*, IEEE Transactions on Information Theory **51** (2005), 4032–4039.
4. G. D. Birkhoff, *Lattice theory*, AMS Colloq. **25** (1967).
5. M. Blum, M. Luby, and R. Rubinfeld, *Self-testing/correcting with applications to numerical problems*, J. Computer and System Sciences **47** (1993), 549–595.
6. A. Bogdanov and E. Viola, *Pseudorandom bits for polynomials*, Proceedings of the 48th annual symposium on Foundations of Computer Science (FOCS) (2007), 41–51.
7. E. Candés, J. Romberg, and T. C. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Inf. Theory **52** (2006), no. 2, 489–509.
8. W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, Geom. Func. Anal. **8** (1998), 529–551.
9. ———, *A new proof of Szemerédi's theorem*, Geom. Func. Anal. **11** (2001), 465–588.
10. R. Graham, B. Rothschild, and J. H. Spencer, *Ramsey theory*, John Wiley and Sons, NY, 1980.
11. B. J. Green and T. C. Tao, *Linear equations in primes*, Annals of Math., to appear.
12. ———, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edin. Math. Soc. **51** (2008), no. 1, 73–153.
13. ———, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.
14. ———, *New bounds for Szemerédi's theorem, I: Progressions of length 4 in finite field geometries*, Proc. Lond. Math. Soc. (3) **98** (2009), no. 2, 365–392.
15. S. Lovett, R. Meshulam, and A. Samorodnitsky, *Inverse conjecture for the gowers norm is false*, STOC '08: Proceedings of the 40th annual ACM symposium on Theory of computing, ACM, New York, NY, 2008, pp. 547–556.

16. A. Samorodnitsky, *Low-degree tests at large distances*, STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, ACM, New York, NY, 2007, pp. 506–515.
17. M. Sudan, L. Trevisan, and S. Vadhan, *Pseudorandom generators without the XOR lemma*, Special issue on the Fourteenth Annual IEEE Conference on Computational Complexity (Atlanta, GA), 1999.
18. T. C. Tao, *Structure and randomness in combinatorics*, Proceedings of the 48th annual symposium on Foundations of Computer Science (FOCS), 2007, pp. 3–18.
19. T. C. Tao and V. H. Vu, *Additive combinatorics*, Cambridge Univ. Press, 2006.

Centre for Mathematical Sciences
Wilberforce Road, Cambridge CB3 0WA, England
E-mail address: b.j.green@dpmms.cam.ac.uk

Department of Mathematics, UCLA,
Los Angeles, CA 90095-1596
E-mail address: tao@math.ucla.edu