Inverse Conjecture for the Gowers norm is false

Shachar Lovett   Roy Meshulam   Alex Samorodnitsky

Abstract

Let p be a fixed prime number, and N be a large integer. The 'Inverse Conjecture for the Gowers norm' states that if the "d-th Gowers norm" of a function $f: \mathbb{F}_p^N \rightarrow \mathbb{F}$ is non-negligible, that is larger than a constant independent of N, then f can be non-trivially approximated by a degree $d-1$ polynomial. The conjecture is known to hold for $d = 2, 3$ and for any prime p. In this paper we show the conjecture to be false for $p = 2$ and for $d = 4$, by presenting an explicit function whose 4-th Gowers norm is non-negligible, but whose correlation any polynomial of degree 3 is exponentially small.

Essentially the same result (with different correlation bounds) was independently obtained by Green and Tao [5]. Their analysis uses a modification of a Ramsey-type argument of Alon and Beigel [1] to show inapproximability of certain functions by low-degree polynomials.

We observe that a combination of our results with the argument of Alon and Beigel implies the inverse conjecture to be false for any prime p, for $d = p^2$.

1 Introduction

We consider multivariate functions over finite fields. The main question of interest here would be whether these functions can be non-trivially approximated by a low-degree polynomial.

Fix a prime number p. Let $\mathbb{F} = \mathbb{F}_p$ be the finite field with p elements. Let $\xi = e^{\frac{2\pi i}{p}}$ be the primitive p-th root of unity. Denote by $e(x)$ the exponential function taking $x \in \mathbb{F}$ to $\xi^x \in \mathbb{C}$. For two functions $f, g: \mathbb{F}^N \rightarrow \mathbb{F}$, let $\langle f, g \rangle := \mathbb{E}_x e(f(x) - g(x))$.

Definition 1.1: A function f is non-trivially approximable by a degree-d polynomial if

$$| \langle f, g \rangle | > \epsilon$$

for some polynomial g of degree at most d in $\mathbb{F}[x_1 ... x_N]$. ∎

More precisely, in this definition we are looking at a sequence f_N of functions and of approximating low-degree polynomials g_N in N variables, and let N grow to infinity. In this paper, the remaining parameters, that is the field size p, the degree d and the offset ϵ are fixed, independent of N.

A counting argument shows that a generic function can not be approximated by a polynomial of low degree. The problems of showing a specific given function to have no non-trivial approximation and of constructing an explicit non-approximable function have been extensively investigated, since solutions to these problems have many applications in complexity (cf. discussion and references in [1, 9, 2]).

This paper studies a technical tool that measures distance from low-degree polynomials. This is the Gowers norm, introduced in [3]. For a function $f : \mathbb{F}^N \rightarrow \mathbb{F}$ and a vector $y \in \mathbb{F}^n$, we take f_y to be the directional derivative of f in direction y by setting

$$f_y(x) = f(x+y) - f(x)$$

For a k-tuple of vectors $y_1...y_k$ we take the iterated derivative in these directions to be

$$f_{y_1...y_k} = \left(f_{y_1...y_{k-1}}\right)_{y_k}$$

It is easy to see that this definition does not depend on the ordering of $y_1...y_k$.

The k-th Gowers "norm" $\|f\|_{U^k}$ of f is

$$\left(\mathbb{E}_{x,y_1...y_k}\left[e\left(f_{y_1...y_k}(x)\right)\right]\right)^{1/2^k}$$

More accurately, as shown in [3], this is indeed a norm of the associated complex-valued function $e(f)$ (for $k \geq 2$).

It is easy to see that $\|f\|_{U^{d+1}}$ is 1 iff f is a polynomial of degree at most d. This is just another way of saying that all order-$(d+1)$ iterative derivatives of f are zero if and only if f is a polynomial of degree at most d. It is also possible to see [4] that $|\langle f, g\rangle| > \epsilon$ for g of degree at most d, implies $\|f\|_{U^{d+1}} > \epsilon$. That is to say, if f is non-trivially close to a degree-d polynomial, this can be detectable via an appropriate Gowers norm.

This discussion naturally leads to the inverse conjecture [4, 7, 8], that is if $(d+1)$-th Gowers norm of f is non-trivial, then f is non-trivially approximable by a degree-d polynomial. This conjecture is easily seen to hold for $d = 1$ and has been proved also for $d = 2$ [4, 7]. It is of interest to prove this conjecture for higher values of d.

In this paper we show this conjecture, which we will refer to as the 'Inverse Conjecture for the Gowers norm', or, informally, as ICGN, to be false. Let S_n be the elementary symmetric polynomial of degree n in N variables, that is

$$S_n(x) = \sum_{S \subseteq [N],\ |S|=n} \prod_{i \in S} x_i$$

We prove two claims about symmetric polynomials. Note that here and below a constant is *absolute* if it does not depend on N.

First, we show Gowers norms of some symmetric polynomials to be non-trivial.

Theorem 1.2: *There is an absolute positive constant ϵ such that for any prime p*

$$\|S_{2p}\|_{U^{p+2}} > \epsilon,$$

Here S_{2p} is viewed as a function over $\mathbb{F} = \mathbb{F}_p$.

Two versions of this result will be useful later.

- A special case $p = 2$.

$$\|S_4\|_{U^4} > \epsilon \tag{1}$$

- An easy generalization: for any $n \geq 2p$,

$$\|S_n\|_{U^{n-p+2}} > \epsilon \tag{2}$$

In the second claim we show a specific symmetric polynomial to have no non-trivial approximation by polynomials of lower degree.

Theorem 1.3: *Let $p = 2$. For any polynomial g of degree 3 holds*

$$|\langle S_4, g\rangle| < \exp\{-\alpha N\} \tag{3}$$

We conjecture the second claim of the theorem to be true for any prime number p, replacing 3 with $p+1$ and 4 with $2p$.

The combination of (1) and (3) shows ICGN to be false for $p = 2$ and $d = 4$.

1.1 Related work

Our results have a large overlap with a recent work of Green and Tao [5].

The paper of Green and Tao has two parts. In the first part ICGN is shown to be true when f is itself a polynomial of degree less than p and $d < p$. In the second part, the conjecture is shown to be false in general. In particular the symmetric polynomial S_4 is shown to be a counterexample for $p = 2$ and $d = 4$.

To proof of non-approximability of S_4 by lower-degree polynomials in [5] uses a modification of a Ramsey-type argument due to Alon and Beigel [1]. Very briefly, this argument shows that if a function over $\mathbb{F}_2$ has a non-trivial correlation with a multilinear polynomial of degree d, then its restriction to a subcube of smaller dimension has a non-trivial correlation with a symmetric polynomial of degree d. The problem of inapproximability by symmetric polynomials turns out to be easier to analyze.

This argument gives a somewhat weaker bounds for non-inapproximability of S_4, in that it shows $\langle S_4, g\rangle < \log^{-c}(N)$ for any degree-3 polynomial g and for an absolute constant $c > 0$.

On the other hand, this argument is more robust than our inapproximability argument. We observe below that it can be readily extended to the case of general prime p and, combined with (2), show ICGN to be false for all p.

1.2 The case of a general prime field

We briefly observe here that a minor adaptation of the Alon-Beigel argument, together with (2), show the symmetric polynomial S_{p^2} to have a non-negligible $\left(p^2\right)$-nd Gowers norm over $\mathbb{F}_p$ and to have no good approximation by lower-degree polynomials. In that, S_{p^2} provides a counterexample to ICGN for any prime p.

Indeed, by monotonicity of the Gowers norms ([4]), and since $p \geq 2$, a direct implication of (2) gives

$$\|S_{p^2}\|_{U^{p^2}} > \epsilon$$

On the other hand, let g be a polynomial of degree less than p^2 in N variables such that $\left\langle S_{p^2}, g\right\rangle > \epsilon$. Note that the Alon-Beigel argument (as given in [1] and in [5]) does not seem to be immediately applicable in this case, since g does not have to be multilinear. A way around this obstacle, is to observe, via an averaging argument, that there is a copy of an N'-dimensional boolean cube $\{0,1\}^{N'}$, such that restrictions S' and g' of S_{p^2} and of g on this subcube satisfy $\langle S', g'\rangle > \epsilon'$, and N', ϵ' depend linearly on N, ϵ'. Without loss of generality assume the coordinates of the boolean cube to be $\{1...N'\}$ and consider the functions S', g' as functions in variables $x_1, ..., x_{N'}$ (with some fixed assignment of values to variables $x_i,\ i > N'$). Now, $S' = \sum_{i=0}^{p^2} a_i S_i$ is a symmetric polynomial of degree p^2 over $\mathbb{F}^{N'}$, with $a_i = 1$, and g' is a polynomial of a degree smaller than p^2. Our gain is in that now g' can be replaced by a multilinear polynomial coinciding with g' on the boolean cube, and hence having a non-trivial correlation with S' on the boolean cube.

Now, the Alon-Beigel argument can be applied to show that the symmetric polynomial S_{p^2} has a non-trivial correlation with a symmetric polynomial h of a smaller degree over the boolean cube $\{0,1\}^{N'}$ viewed as a subset of $\mathbb{F}^{N'}$. This, however, couldn't be true due to a theorem of Lucas, which implies that for a boolean vector x with Hamming weight $w = \sum_{i=1}^{N'} x_i$, the value $S_{p^2}(x)$ depends only on the 3-rd digit in the representation of w in base p, while the value of h depends only on the first 2 digits.

This completes the argument. We conclude with an observation that this argument directly extends to S_{p^k} for any $k > 1$.

Here is a brief overview of the rest of the paper. Section 2 defines relevant notions and contains proofs of several technical claims. Theorem 1.2 is proved in Section 3. Theorem 1.3 is proved in Section 4.

2 Some useful notions and claims

2.1 Some multilinear polynomials and their properties

In this sub-section we introduce and discuss certain polynomials over the finite field $\mathbb{F}$. These polynomials can be conveniently viewed as multi-linear functions on matrices whose entries are elements of $\mathbb{F}$, or formal variables with values in the field. A basic object we consider is a rectangular $n \times N$ matrix, $N \geq n$. A matrix M with rows $r_1 ... r_n$ will be denoted by $M[r_1 ... r_n]$.

Sometimes there will be repeated rows. In such a case we consider a partition $\lambda = (\lambda_1 ... \lambda_k)$ of $[n]$, that is λ_i are (possibly empty) subsets of $[n]$, whose disjoint union is $[n]$. We denote by $M_\lambda[r_1...r_k]$ the matrix whose rows in positions indexed by elements of λ_i equal r_i. Note that the partition λ is ordered, in that the ordering of the sets λ_i is relevant. We use the notation $\{\lambda_1...\lambda_k\}$ for an unordered partition.

First, we introduce the "symmetric" function $\mathcal{S}$. We define $\mathcal{S}(M)$ to be the sum of all the permanental minors of M, that is

$$\mathcal{S}(M) := \sum_{C \subseteq [N], |C|=n} Per\left(M_C\right),$$

where M_C is an $n \times n$ submatrix of M which is obtained by deleting all the columns of M except these with indices in C.

Let $\lambda = (\lambda_1 ... \lambda_k)$ be a partition of $[n]$, and set $\ell_i = |\lambda_i|$. Clearly $\mathcal{S}\left(M_\lambda\right)$ depends only on the cardinalities ℓ_i of λ_i. This leads to the notation $M\left[r_1^{(\ell_1)}...r_k^{(\ell_k)}\right]$ which denotes the matrix in which the row r_1 appears ℓ_1 times, followed by ℓ_2 appearances of the row r_2 and so on. In this notation, therefore

$$\mathcal{S}\left(M_{(\lambda_1...\lambda_k)}[r_1...r_k]\right) = \mathcal{S}\left(M\left[r_1^{(|\lambda_1|)}...r_k^{(|\lambda_k|)}\right]\right)$$

The second matrix function we consider is the "forward" function $\mathcal{F}$, with

$$\mathcal{F}(M\left[r_1...r_n\right]) = \sum_{C \subseteq [N], |C| = \{j_1 < j_2 < ... < j_n\}} \prod_{i=1}^{n} r_i\left(j_i\right)$$

Here $r_i(j)$ denote the j-th coordinate of the vector r.

To connect the two notions, observe that

$$\mathcal{S}(M[r_1...r_n]) = \sum_{\sigma} \mathcal{F}(M[r_{\sigma_1}...r_{\sigma_n}])$$

where σ runs over all permutations on n items.

The last function we consider is a "hybrid" function $\mathcal{H}$ which has some 'symmetric' and some 'forward' properties. Let $\lambda = (\lambda_1...\lambda_k)$ be an ordered partition of $[n]$ with k terms. For another such partition $\theta = (\theta_1...\theta_k)$ of $[n]$ write $\theta \sim \lambda$ if $|\theta_1| = |\lambda_1|$,...,$|\theta_k| = |\lambda_k|$. We define

$$\mathcal{H}\left(M_\lambda[r_1...r_k]\right) = \sum_{C \subseteq [N], |C| = \{j_1 < j_2 < ... < j_n\}} \sum_{\theta \sim \lambda} \prod_{t=1}^{k} \prod_{i \in \theta_t} r_t\left(j_i\right)$$

An alternative view of the functions $\mathcal{S}, \mathcal{F}$ and $\mathcal{H}$ might be helpful at this point. Consider the set of *paths* which are one-to-one functions from $[n]$ to $[N]$. Let us call a path ρ monotone on a subset $\{i_1 < i_2 < ... < i_\ell\}$ of $[n]$ if $\rho\left(i_1\right) < \rho\left(i_2\right) < ... < \rho\left(i_\ell\right)$. A path is (fully) monotone if it is monotone on $[n]$. Then, for a partition $\lambda = (\lambda_1...\lambda_k)$ of $[n]$ and an $n \times N$ matrix $M = M_\lambda$,

$$\mathcal{S}(M) = \sum_{\text{all } \rho} \prod_{i=1}^{n} M_{i,\rho(i)}$$

$$\mathcal{F}(M) = \sum_{\text{monotone } \rho} \prod_{i=1}^{n} M_{i,\rho(i)}$$

$$\mathcal{H}(M) = \sum_{\rho \text{ monotone on } \lambda_1 ... \lambda_k} \prod_{i=1}^{n} M_{i,\rho(i)}$$

Note that for the function $\mathcal{H}$, similarly to the symmetric function $\mathcal{S}$, holds

$$\mathcal{H}\left(M_{(\lambda_1 ... \lambda_k)}[r_1 ... r_k]\right) = \mathcal{H}\left(M\left[r_1^{(|\lambda_1|)} ... r_k^{(|\lambda_k|)}\right]\right)$$

Observe also that if $\lambda = (\{1\} ... \{n\})$ then $\mathcal{S}(M) = \mathcal{H}(M)$. If $\lambda = (\{[n]\})$ then $\mathcal{F}(M) = \mathcal{H}(M)$ and $\mathcal{S}(M) = n! \cdot \mathcal{F}(M) = n! \cdot \mathcal{H}(M)$. For a general $\lambda = (\lambda_0 ... \lambda_k)$

$$\mathcal{S}(M) = \left(\prod_{t=1}^{k} |\lambda_t|!\right) \cdot \mathcal{H}(M) \tag{4}$$

Note that this is an identity in $\mathbb{F}$. In particular, if one of the terms λ_i has cardinality at least p then $\mathcal{S}(M) = 0$ and (4) provides no information.

To simplify the notation we will usually write $\mathcal{S}(r_1 ... r_n)$ for $\mathcal{S}(M[r_1 ... r_n])$, $\mathcal{F}_\lambda(r_1 ... r_k)$ for $\mathcal{F}\left(M_\lambda[r_1 ... r_k]\right)$ and so on.

2.2 Directional derivatives of symmetric polynomials

The functions we have defined are relevant to the discussion here for two reasons. First, the elementary symmetric polynomial $S_n(x)$ in N variables can be viewed as the forward function $\mathcal{F}$ applied to the matrix $M[x ... x]$, where M has n identical rows equal to x. In our notation,

$$S_n(x) = \mathcal{F}_{\{[n]\}}(x)$$

Second, it is possible to write a directional derivative $(S_n)_{y_1 ... y_k}$ of S_n of any order as a combination of values of $\mathcal{F}$ on explicitly defined matrices M whose rows are either the indeterminate x or the directions y_i.

The basic observation here is the following lemma which is straightforward from the definition of directional derivative.

Lemma 2.1: *Let a polynomial $P(x)$ in N variables be given by*

$$P(x) = \mathcal{F}_{(\lambda_0 ... \lambda_k)}(x, y_1 ... y_k)$$

Then

$$P_z(x) = \sum_{A \subset \lambda_0} \mathcal{F}_{(A, \lambda_0 \setminus A, \lambda_1 ... \lambda_k)}(x, z, y_1 ... y_k)$$

In words, when we take the derivative of such a polynomial in direction z, we replace some of the rows which contained x with z.

As a corollary we have a following expression for higher order derivatives of a symmetric polynomial.

Proposition 2.2: *Let $k \leq n$, then*

$$(S_n)_{y_1 \ldots y_k}(x) = \sum_{m=0}^{n-k} \sum_{\ell_1 \ldots \ell_k \geq 1,\ \sum_i \ell_i = n-m} \mathcal{H}\left(x^{(m)}, y_1^{(\ell_1)} \ldots y_k^{(\ell_k)}\right)$$

Proof: Iterating Lemma 2.1,

$$(S_n)_{y_1 \ldots y_k}(x) = \sum_{\lambda = (\lambda_0, \lambda_1 \ldots \lambda_k)} \mathcal{F}_\lambda(x, y_1 \ldots y_k)$$

where the summation is over partitions λ such that λ_i are not empty for $i = 1 \ldots k$. Rearranging, this is

$$\sum_{m=0}^{n-k} \sum_{\ell_1 \ldots \ell_k \geq 1,\ \sum_i \ell_i = n-m} \sum_{\lambda:\ |\lambda_0| = m, |\lambda_1| = \ell_1 \ldots |\lambda_k| = \ell_k} \mathcal{F}_\lambda(x, y_1 \ldots y_k) =$$

$$\sum_{m=0}^{n-k} \sum_{\ell_1 \ldots \ell_k \geq 1,\ \sum_i \ell_i = n-m} \mathcal{H}\left(x^{(m)}, y_1^{(\ell_1)} \ldots y_k^{(\ell_k)}\right)$$

∎

We can give explicit expressions for the coefficients of $(S_n)_{y_1 \ldots y_k}(x)$. Fix m indices $j_1 < j_2 < \ldots < j_m$ for $0 \leq m \leq n-k$, and let a be the coefficient of $x_{j_1} \cdots x_{j_m}$ in $(S_n)_{y_1 \ldots y_k}$.

Corollary 2.3:

- $$a = \sum_{\ell_1 \ldots \ell_k \geq 1,\ \sum_i \ell_i = n-m} \mathcal{H}^{\{j_1 \ldots j_m\}}\left(y_1^{(\ell_1)} \ldots y_k^{(\ell_k)}\right)$$

- *If $k + m + p > n + 1$ then*

$$a = \sum_{\ell_1 \ldots \ell_k \geq 1,\ \sum_i \ell_i = n-m} \left(\prod_{i=1}^{k} \ell_i!\right)^{-1} \cdot \mathcal{S}^{\{j_1 \ldots j_m\}}\left(y_1^{(\ell_1)} \ldots y_k^{(\ell_k)}\right)$$

Here, for a subset of indices $T \subseteq [N]$, $\mathcal{H}^T(M)$ returns the value of the matrix function $\mathcal{H}$ applied to the $n \times (N - |T|)$ matrix obtained from M by deleting columns in T. The function $\mathcal{S}^T(M)$ is defined similarly.

Proof: The first claim is immediate from Proposition 2.2. The second claim follows from the first claim, from (4), and from the simple observation that if $k + m + p > n + 1$ then $\ell_i < p$ for $i = 1 \ldots k$ in the above summation, which means $\ell_i!$ is invertible in $\mathbb{F}_p$. ∎

Example 2.4: The following "toy" example will be relevant for the case of the binary field. It is sufficiently simple to illustrate what's going on behind the cumbersome formulas. Consider $P = (S_4)_{y,z}$. Then P is a quadratic polynomial and for $1 \leq i < j \leq N$

$$\mathrm{coef}_{x(i)x(j)}(P) = \sum_{k \neq l,\ k,l \notin \{i,j\}} y(k)z(l) = \mathcal{S}^{\{i,j\}}(y,z)$$

∎

Continuing with the same example, note that it convenient to express the symmetric function $\mathcal{S}(y,z)$ via inner products of vectors $y, z, \mathbf{1}$, where $\mathbf{1}$ is the all-1 vector of length N.

$$\mathcal{S}(y,z) = \sum_{k \neq l} y(k)z(l) = \langle y, \mathbf{1} \rangle \cdot \langle z, \mathbf{1} \rangle - \langle yz, \mathbf{1} \rangle$$

Here we take yz to be the vector whose coordinates are point-wise inner products of the coordinates of y and z, that is $(yz)(i) = y(i)z(i)$. Of course, $\langle yz, \mathbf{1} \rangle$ is the same as $\langle y, z \rangle$.

Similarly, we can express the 'incomplete' symmetric function $\mathcal{S}^{\{i,j\}}(y,z)$ via the complete symmetric function $\mathcal{S}(y,z)$ minus forbidden terms, as follows

$$\mathcal{S}^{\{i,j\}}(y,z) = \mathcal{S}(y,z) - \Big(z(i) + z(j)\Big) \langle y, \mathbf{1} \rangle - \Big(y(i) + y(j)\Big) \langle z, \mathbf{1} \rangle + \Big(y(i)z(j) + y(j)z(i)\Big)$$

Note the "inclusion-exclusion" structure in the two expressions above. (To make it even clearer we use "+" and "-" notation, though in the binary field both are, of course, the same.) This structure becomes more evident as we pass to our next order of business, which is expressing, for general n and k, the coefficients of $(S_n)_{y_1 \ldots y_k}$ via inner products of vectors $y_1 \ldots y_k, \mathbf{1}$.

2.3 Inclusion-Exclusion formulas for symmetric functions

Some notation: Given m vectors $y_1 \ldots y_m$ and a subset $\tau \subseteq [m]$, let y_τ to be vector whose coordinates are point-wise products of the corresponding coordinates of y_i, $i \in \tau$. Let $\mathcal{S}(y[\tau])$ for the value of the function $\mathcal{S}$ on a matrix with $|\tau|$ rows y_i, $i \in \tau$. Let $\langle y_\tau \rangle$ be the polynomial $\langle y_\tau, \mathbf{1} \rangle = \sum_{j=1}^{N} \prod_{i \in \tau} y_i(j)$.

We start with an auxiliary lemma expressing the incomplete symmetric function $\mathcal{S}^{\{k\}}(r_1 \ldots r_n)$ as a polynomial in the k-th coordinate of the vectors r_i and in complete symmetric functions applied to sub-matrices of $M[r_1 \ldots r_n]$.

Lemma 2.5:

$$\mathcal{S}^{\{k\}}(r_1 \ldots r_n) = \sum_{\tau \subseteq [n]} (-1)^{|\tau|} (|\tau|)! \cdot r_\tau(k) \cdot \mathcal{S}\left(r\Big[[n] \setminus \tau\Big]\right)$$

From now on we assume $r_\emptyset$ to be the all-1 vector, and $\mathcal{S}(r[\emptyset])$ to equal 1.

Proof: The proof is by induction on n. For $n = 1$ both sides equal $\sum_{j=1}^{N} r_1(j) - r_1(k)$.

For $n > 1$, observe that

$$\mathcal{S}^{\{k\}}\left(r_1 ... r_n\right) = \mathcal{S}\left(r_1 ... r_n\right) - \sum_{i=1}^{n} r_i(k) \cdot \mathcal{S}^{\{k\}}\left(r\Big[[n] \setminus \{i\}\Big]\right)$$

and the claim is easily verified using the induction hypothesis. ∎

Now we can state two main claims of this section. The first expresses the complete symmetric function $\mathcal{S}\left(r_1 ... r_n\right)$ via inner products $\langle r_T \rangle$.

Proposition 2.6:

$$\mathcal{S}\left(r_1 ... r_n\right) = \sum_{\lambda=\{\lambda_1 ... \lambda_m\}} \prod_{t=1}^{m} \left((-1)^{|\lambda_t|-1} \left(|\lambda_t| - 1\right)! \cdot \langle r_{\lambda_t} \rangle\right)$$

In this summation $\lambda = \{\lambda_1 ... \lambda_m\}$ runs over all unordered partitions of $[n]$ with non-empty λ_i.

Proof: Again, the proof is by induction on n. For $n = 1$ both sides equal $\sum_{j=1}^{N} r_1(j)$. For $n > 1$ we have

$$\mathcal{S}(r_1 ... r_n) = \sum_{k=1}^{N} r_n(k) \cdot \mathcal{S}^{\{k\}}(r_1 ... r_{n-1})$$

Using Lemma 2.5 and the induction hypothesis,

$$\mathcal{S}(r_1 ... r_n) = \sum_{k=1}^{N} r_n(k) \cdot \sum_{\tau \subseteq [n-1]} (-1)^{|\tau|}(|\tau|)! \cdot r_\tau(k) \cdot \mathcal{S}\left(r\Big[[n-1] \setminus \tau\Big]\right) =$$

$$\sum_{\tau \subseteq [n-1]} (-1)^{|\tau|}(|\tau|)! \cdot \left\langle r_{\tau \cup [n]} \right\rangle \cdot \mathcal{S}\left(r\Big[[n-1] \setminus \tau\Big]\right)$$

Consider the summand corresponding to $\tau = [n-1]$. Recall the boundary assumption $\mathcal{S}\left(r[\emptyset]\right) = 1$. Hence this summand is $(-1)^{n-1}(n-1)! \cdot \left\langle r_{[n]} \right\rangle$. This summand therefore corresponds to the partition $\lambda = \{[n]\}$ in the claim of the proposition.

For τ a proper subset of $[n-1]$, we use the induction hypothesis to obtain

$$\mathcal{S}(r_1 ... r_n) = \sum_{\tau \subseteq [n-1]} (-1)^{|\tau|}(|\tau|)! \cdot \left\langle r_{\tau \cup [n]} \right\rangle \cdot \sum_{\theta=\{\theta_1 ... \theta_l\}} \prod_{t=1}^{l} \left((-1)^{|\theta_t|-1} \left(|\theta_t| - 1\right)! \cdot \langle r_{\theta_t} \rangle\right) +$$

$$(-1)^{n-1}(n-1)! \cdot \left\langle r_{[n]} \right\rangle$$

Here θ runs over all the unordered partitions of $[n-1] \setminus \tau$ with non-empty θ_i. Observe that each pair (τ, θ) leads to a unique partition $\lambda = \{\lambda_1 ... \lambda_{l+1}\} = \{\theta_1 ... \theta_l, \tau \cup [n]\}$ of $[n]$. Rearranging the terms, the last summation can be written as

$$\sum_{\lambda=(\lambda_1 ... \lambda_m)} \prod_{t=1}^{m} \left((-1)^{|\lambda_t|-1} \left(|\lambda_t| - 1\right)! \cdot \langle r_{\lambda_t} \rangle\right)$$

completing the proof of the proposition. ∎

The second claim expresses the incomplete symmetric function $\mathcal{S}^{\{j_1...j_k\}}(r_1...r_n)$ as a polynomial in the missing coordinates $j_1...j_k$ of the vectors r_i and in complete symmetric functions applied to sub-matrices of $M[r_1...r_n]$. Note that Lemma 2.5 is a special case $k=1$ of this claim.

Proposition 2.7:

$$\mathcal{S}^{\{j_1...j_k\}}(r_1...r_n) = \sum_{\tau=(\tau_1...\tau_k)} \prod_{t=1}^{k} \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t)\right) \cdot \mathcal{S}\left(r\left[[n] \setminus \cup_t \tau_t\right]\right)$$

Here the summation is on all ordered set systems τ such that the terms τ_t are disjoint subsets of $[n]$. The terms may also be empty.

Proof: The proof is by induction on k and n. The case $k=1$ is treated in Lemma 2.5.

Consider the case $n=1$. On one hand $\mathcal{S}^{\{j_1...j_k\}}(r_1) = \sum_{j=1}^{N} r_1(j) - \sum_{t=1}^{k} r_1(j_t)$. We claim that this value can be also represented as

$$\sum_{\tau=(\tau_1...\tau_k)} \prod_{t=1}^{k} \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t)\right) \cdot \mathcal{S}\left(r\left[[1] \setminus \cup_t \tau_t\right]\right)$$

Here τ_i are disjoint subsets of $[1]$. Observe that there are $k+1$ summands in this expression, corresponding to different set systems τ. Let $\tau^{(0)}$ denote the set system with k empty terms, and let $\tau^{(t)}$, for $t=1...k$ denote the set system with $\tau_t = \{1\}$ and all the remaining terms are empty. The summand corresponding to $\tau^{(0)}$ is $\mathcal{S}(r_1) = \sum_{j=1}^{N} r_1(j)$. The summand corresponding to $\tau^{(t)}$ is $(-r_1(j_t)) \cdot \mathcal{S}(r_\emptyset) = -r_1(j_t)$, and we are done in this case.

For $k, n > 1$, we have

$$\mathcal{S}^{\{j_1...j_k\}}(r_1...r_n) = \mathcal{S}^{\{j_1...j_{k-1}\}}(r_1...r_n) - \sum_{i=1}^{n} r_i(j_k) \cdot \mathcal{S}^{\{j_1...j_k\}}\left(r\left[[n] \setminus \{i\}\right]\right)$$

By the induction hypothesis, this is

$$\sum_{\theta=(\theta_1...\theta_{k-1})} \prod_{t=1}^{k-1} \left((-1)^{|\theta_t|}(|\theta_t|)! \cdot r_{\theta_t}(j_t)\right) \cdot \mathcal{S}\left(r\left[[n] \setminus \cup_t \theta_t\right]\right) -$$

$$\sum_{i=1}^{n} r_i(j_k) \cdot \sum_{\mu^{(i)}=\left(\mu_1^{(i)}...\mu_k^{(i)}\right)} \prod_{u=1}^{k} \left((-1)^{|\mu_u^{(i)}|}(|\mu_u^{(i)}|)! \cdot r_{\mu_u^{(i)}}(j_u)\right) \cdot \mathcal{S}\left(r\left[[n] \setminus \cup_t \mu_t^{(i)} \setminus \{i\}\right]\right)$$

Here the summation is on all ordered set systems θ such that the terms θ_t are disjoint subsets of $[n]$ and on ordered set systems $\mu^{(i)}$, $i=1...n$ such that the terms $\mu_u^{(i)}$ are disjoint subsets of $[n] \setminus \{i\}$.

Given a set system $\theta = (\theta_1 ... \theta_{k-1})$ we define a set system $\tau = (\tau_1 ... \tau_k)$ by setting $\tau_t = \theta_t$, $t = 1...k-1$ and $\tau_k = \emptyset$. Given a set system $\mu^{(i)} = \left(\mu_1^{(i)} ... \mu_k^{(i)}\right)$ we define a set system $\tau = (T_1 ... \tau_k)$ by setting $\tau_u = \mu_u^{(i)}$, $u = 1...k-1$ and $\tau_k = \mu_k^{(i)} \cup \{i\}$. In both cases we have obtained a set system of the type we want, that is an ordered family of k disjoint subsets of $[n]$. Moreover, each such system with empty k-th term is obtained exactly once, from the corresponding θ-system, and each system with non-empty k-th term τ_k is obtained exactly $|\tau_k|$ times, from systems $\mu^{(i)}$ with $i \in \tau_k$. Rearranging the terms and the signs, the last expression is precisely

$$\sum_{\tau=(\tau_1...\tau_k)} \prod_{t=1}^{k} \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t)\right) \cdot \mathcal{S}\left(r\Big[[n] \setminus \cup_t \tau_t\Big]\right),$$

completing the proof. ∎

2.4 Some properties of Gowers' norms

The main result in this subsection shows that if a function from $\mathbb{F}^N$ to $\mathbb{F}$ is fixed on a subset of $\mathbb{F}^N$ defined by low-degree polynomial constraints, then it has a non-trivial Gowers norm of an appropriate order.

Recall that for a vector $x \in \mathbb{F}^N$, x^i stands for a vector in $\mathbb{F}^N$ whose coordinates are i-th powers of the coordinates of x.

Proposition 2.8: *Let K be an absolute constant. Let $y_{i,j}$, $i = 1...p-1$, $j = 1...K$, be $K(p-1)$ vectors in $\mathbb{F}^N$. Let M be a subset of $\mathbb{F}^N$ defined by the constraints $\langle x^i, y_{i,j}\rangle = 0$ for all i, j.*

Let f be a function from $\mathbb{F}^N$ to $\mathbb{F}$. Assume that f is fixed on M. Then

$$\|f\|_{U^p} > \left(\frac{|M|}{2^N}\right)^2 =: Pr^2\{M\}$$

Proof: Let $f_{|M} \equiv c_0$.

Consider a subspace V of polynomials of degree at most $p-1$ in $\mathbb{F}[x_1...x_N]$ spanned by the polynomials $\langle x^i, y_{i,j}\rangle$, for all i, j. We will first find a polynomial $g \in V$ such that $|\langle f, g\rangle| \geq Pr\{M\}$. This, combined with a lemma from [4], will imply the claim of the proposition.

Let $\mathbf{b} = (b_{i,j})$, $i = 1...p-1$, $j = 1...K$, be a matrix with entries in $\mathbb{F}$. Let $c \in \mathbb{F}$. Set

$$\mu(\mathbf{b}, c) = Pr\Big\{x : \ f(x) = c \ \wedge \ \langle x^i, y_{i,j}\rangle = b_{i,j} \text{ for all } i, j\Big\}$$

Note that, by assumption, for a zero matrix $\mathbf{b}$ holds $\mu(\mathbf{b}, c_0) = Pr\{M\}$. In other words, $\mu(\mathbf{b}, c) = 0$ and for $\mathbf{b} = 0$ any $c \neq c_0$.

Now, for any $g(x) = \sum_{i,j} a_{i,j} \langle x^i, y_{i,j}\rangle$ in V holds

$$\langle f, g\rangle = \mathbb{E}e(f-g) = \sum_{\mathbf{b},c} \mu(\mathbf{b}, c) \cdot e\left(c - \langle \mathbf{a}, \mathbf{b}\rangle\right)$$

where $\mathbf{a} = (a_{i,j})_{i,j}$ and $\langle \mathbf{a}, \mathbf{b} \rangle = \sum_{i,j} a_{i,j} b_{i,j}$. Averaging over V, we have

$$\mathbb{E}_{g \in V} \langle f, g \rangle = \frac{1}{|V|} \sum_{\mathbf{a}} \sum_{\mathbf{b},c} \mu(\mathbf{b}, c) \cdot e\left(c - \langle \mathbf{a}, \mathbf{b} \rangle\right) = \frac{1}{|V|} \sum_{\mathbf{b},c} \mu(\mathbf{b}, c) \cdot e(c) \sum_{\mathbf{a}} e\left(- \langle \mathbf{a}, \mathbf{b} \rangle\right) =$$

$$\sum_{c} \mu(0, c) \cdot e(c) = \mu\left(0, c_0\right) \cdot e\left(c_0\right) = Pr\{M\} \cdot e\left(c_0\right)$$

This means, there is $g \in V$ with $| \langle f, g \rangle | \geq Pr\{M\}$. We conclude the proof of the proposition by quoting a lemma from [4], which states that $| \langle f, g \rangle | \geq \epsilon$ implies $\|f\|_{U^p} \geq \epsilon$. ∎

2.5 Asymptotic uniformity and independence of some random variables

In this subsection we deal with another property of multiviarite polynomials. Let n be fixed integer and let N be an integer parameter growing to infinity. Let $r_1 ... r_n$ be n vectors in $\mathbb{F}^N$. Let $\kappa = (k_1 ... k_n)$ be a non-zero sequence of integers $0 \leq k_i < p$. For each such sequence define a polynomial $X_\kappa\left(r_1, ..., r_n\right) = \sum_{j=1}^N \prod_{i=1}^n r_i^{k_i}(j)$.

Now,let $r_1 ... r_n$ be chosen uniformly and independently from $\mathbb{F}^N$. We claim that for a large N the random variables $X_\kappa\left(r_1, ..., r_n\right)$ are nearly independent and uniformly distributed over $\mathbb{F}$. Let $X = \left(X_\kappa\right)_\kappa$, and let $K = p^n$.

Proposition 2.9: *Let U be the uniform distribution on $\mathbb{F}^K$. Let P be distribution of X on $\mathbb{F}^K$. Let $\| \cdot \|$ denote the statistical (l_1) distance between distributions.*

Then there is a constant $c > 0$ depending on n, p but not on N such that

$$\|P - U\| \leq \exp\{-cN\}$$

Proof: We start from a simple observation that Fourier transform of a uniform distribution is the delta function at 0. In addition, the two following statements are equivalent up to constants: 'a distribution is exponentially close to uniform' and 'all non-zero Fourier coefficients of the distribution are exponentially close to zero'. Accordingly, we will show that all the non-zero Fourier coefficients of P tend exponentially fast in N to zero.

Consider a character $\chi(y) = \xi^{\langle y, a \rangle}$, corresponding to a non-zero vector $a = \left(a_\kappa\right)_\kappa \in \mathbb{F}^K$. (Recall that $\xi = e^{2\pi i/p}$ is the p-th primitive root of unity.) Then, normalizing appropriately,

$$\widehat{P}(\chi) = \sum_y P(y) \bar{\chi}(y) = \sum_y Pr\{X = y\} \cdot \xi^{-\sum_\kappa a_\kappa y_\kappa} = \mathbb{E} \xi^{-\sum_\kappa a_\kappa X_\kappa}$$

Let P_a denote the distribution of the random variable $X_a = \sum_\kappa a_\kappa X_\kappa$. Then we have shown $\widehat{P}(\chi) = \widehat{P_a}(1)$. We will show the non-zero Fourier coefficients of P_a to be exponentially small, completing the proof of the proposition.

We have

$$X_a\left(r_1, ..., r_n\right) = \sum_\kappa a_\kappa P_\kappa\left(r_1, ..., r_n\right) = \sum_{j=1}^N \sum_{\kappa = (k_1 ... k_n)} a_\kappa \prod_{i=1}^n r_i^{k_i}(j)$$

Let x_i be elements of the field $\mathbb{F}$. Consider an n-variate polynomial

$$Q(x_1...x_n) = \sum_{\kappa=(k_1...k_n)} a_\kappa \prod_{i=1}^{n} x_i^{k_i}$$

Since not all of the coefficients a_κ are zero, and since all κ are non-zero sequences, Q is a multi-variate polynomial of degree at least 1 in $\mathbb{F}[x_1...x_n]$, and therefore attains at least two values with probability bounded away from zero. Now, $X_a = \sum_{j=1}^{N} Q\left(r_1(j)...r_n(j)\right)$ is a sum of N independent copies of Q. Let μ denote the distribution of Q on $\mathbb{F}$. Then the distribution P_a of X_a is μ^{*N}, the N-wise convolution of μ with itself. Since p is prime, $\widehat{\mu}(0) = 1$, and $|\widehat{\mu}| < 1$ everywhere else. Therefore, $\widehat{P_a} = (\widehat{\mu})^N$ tends to the delta function at 0 exponentially fast in N, completing the proof. ∎

2.6 Estimates on the number of common zeroes of some families of polynomials

The main claim of this subsection is the following proposition.

Proposition 2.10: *Let M be the ring of $\mathbb{F}$-valued functions on $\mathbb{F}^N$, that is $M = \mathbb{F}[x_1...x_N]/I$, where I is the ideal $\left(x_1^p - x, ..., x_N^p - x\right)$. Let $f_1...f_K$ be polynomials in M. Let S be the set of common zeroes of $f_1...f_K$, that is*

$$S = \left\{u \in \mathbb{F}^N : \; f_1(u) = ... = f_K(u) = 0\right\}$$

Then

$$|S| \le dim\,(M/J)$$

where J is the ideal generated by $\{f_i\}$, and $dim\,(M/J)$ denotes the dimension of $dim\,(M/J)$, viewed as a vector space over $\mathbb{F}$.

Proof: For each $u \in S$, let $q_u \in M$ be defined by $q_u(u) = 1$ and $q_u(v) = 0$ for all $v \neq u$. We will show that the family $\{q_u + J\}_{u \in S}$ is linearly independent in M/J. This will immediately imply the claim of the proposition.

Consider a linear combination $q = \sum_{u \in S} \lambda_u q_u$ such that $q \in J$. Let $v \in S$. We compute $q(u)$ in two ways. First, since $q \in J$, we have $q(v) = 0$. On the other hand, $q(v) = \sum_{u \in S} \lambda_u q_u(v) = \lambda_v$. This shows $\lambda_v = 0$ for all $v \in S$, completing the proof. ∎

In some cases, the dimension of M/J is easy to estimate.

Lemma 2.11: *Let $p = 2$, let $K = \binom{N}{k}$, and let $\{f_I\}$ be indexed by k-subsets I of $[N]$. Assume that for any such subset I holds*

$$deg\left(f_I(x) - \prod_{i \in I} x_i\right) \le k - 1 \tag{5}$$

Then,

$$dim\,(M/J) \le \sum_{j=0}^{k-1} \binom{N}{j}$$

Proof: We will construct a generating subset of the vector space M/J of cardinality at most $\sum_{j=0}^{k-1}\binom{N}{j}$. We start from a trivial generating set $\{m+J\}$, where m runs through all the 2^N multi-linear monomials in N variables. Now, in the factor space M/J, we can replace any product of k variables, $\prod_{i\in I} x_i$, by a polynomial of degree smaller than k. Iterating this procedure, we arrive to a generating set spanned by $\{s+J\}$, where s now runs through $\sum_{j=0}^{k-1}\binom{N}{j}$ monomials of degree at most $k-1$. ∎

3 Proof of Theorem 1.2

We need to show that

$$\|S_{2p}\|_{U^{p+2}} > \epsilon$$

for an absolute constant ϵ.

We remark that (2) can be shown exactly in the same way, replacing $2p$ with n and $p+2$ with $n-p+2$ throughout.

Recall ([4]) that $\|f\|_{U^{p+2}} = \mathbb{E}_{y,z}^{1/2^{p+2}} \|f_{y,z}\|_{U^p}^{2^p}$. Since the Gowers' norms are nonnegative, it will suffice to show that $\|f_{y,z}\|_{U^p}$ is non-negligible for a non-negligible fraction of directions y,z.

Let

$$A = \left\{(y,z): \ \left\langle y^a, z^b \right\rangle = 0 \text{ for all } 0 \le a,b < p\right\}$$

By Proposition 2.9, for uniformly and independently chosen directions y,z, and for a sufficiently large N, the probability of A is very close to p^{-p^2}. Therefore, A is a non-negligible event. We will now show that for any $(y,z)\in A$ holds $\|f_{y,z}\|_{U^p} > \epsilon'(y,z)$, for an appropriate function ϵ'.

Fix (y,z) in A. Let $f = (S_{2p})_{y,z}$. Let

$$M = M(y,z) = \left\{x: \ \left\langle x^i, y^a z^b \right\rangle = 0 \text{ for all } 1 \le i \le p-1, 0 \le a,b < p\right\}$$

We will show that f is fixed on M. Assuming this, by Proposition 2.8, we have $\|f_{y,z}\|_{U^p} > Pr^2\{M\}$, and therefore

$$\|f\|_{U^{p+2}}^{2^{p+2}} = \mathbb{E}_{y,z}\|f_{y,z}\|_{U^p}^{2^p} \ge Pr\{A\}\cdot \mathbb{E}_{(y,z)\in A} Pr^{2^{p+1}}\{M(y,z)\} \ge$$

$$Pr\{A\}\cdot \mathbb{E}_{(y,z)\in A}^{2^{p+1}} Pr\{M(y,z)\} \ge \left(Pr\{A\}\cdot \mathbb{E}_{(y,z)\in A} Pr\{M(y,z)\}\right)^{2^{p+1}} =$$

$$Pr^{2^{p+1}}\left\{x: \ \left\langle x^i y^a z^b\right\rangle = 0 \text{ for all } 0 \le a,b,i \le p-1\right\} \ge \Omega\left(p^{-p^3\cdot 2^{p+1}}\right)$$

The last inequality follows from Proposition 2.9, since random variables $\left\langle x^i y^a z^b\right\rangle$ are asymptotically uniform and independent.

It remains to prove the following fact.

Lemma 3.1: *Let x, y, z be three vectors in $\mathbb{F}^N$ satisfying $\langle x^i y^a z^b \rangle = 0$ for all $0 \le a, b, i \le p-1$. Then*

$$(S_{2p})_{y,z}(x) = \mathcal{H}\left(y^{(p)}, z^{(p)}\right)$$

Proof: By Proposition 2.2,

$$(S_{2p})_{y,z}(x) = \sum_{m=0}^{2p-2} \sum_{a,b \ge 1,\ a+b=2p-m} \mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$$

We claim that all of the summands on the right, except (possibly) $\mathcal{H}\left(y^{(p)}, z^{(p)}\right)$ are 0.

There are two possible cases to consider. The easier case is when $a, b, m < p$. In such a case, by (4), $\mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$ is proportional to $\mathcal{S}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$. By Proposition 2.6, the symmetric function $\mathcal{S}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$ is a polynomial in $\langle x^i y^a z^b \rangle$, which vanishes when all of these inner products are 0.

In the second case, one of the indices a, b, m is at least p. Note, that there could be at most one such index (barring the case $a = b = p$). We may assume this index is m. We claim that in this case $\mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$ can be written as a linear combination of hybrid functions $\mathcal{H}\left(x^{(\ell)}, r_1, ..., r_{m-\ell}\right)$, where $\ell < m$ and the vectors r_i are of the form $x^\alpha y^\beta z^\gamma$. Note that this will suffice to prove the lemma, since iterating this step will express $\mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$ as a linear combination of symmetric functions in r_i, and these functions vanish.

Consider $\mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right)$. For notational convenience, let $w_1 ... w_{a+b}$ stand for the vectors $y...y, z...z$ (y taken a times and z taken b times). Note that both a and b are smaller than p. Using Corollary 2.3 and Proposition 2.7,

$$\mathcal{H}\left(x^{(m)}, y^{(a)}, z^{(b)}\right) = (a! \cdot b!)^{-1} \cdot \sum_{i_1 < i_2 < ... < i_m} x_{i_1} x_{i_2} \cdots x_{i_m} \mathcal{S}^{\{i_1 ... i_m\}}\left(y^{(a)}, z^{(b)}\right) =$$

$$(a! \cdot b!)^{-1} \cdot \sum_{i_1 < i_2 < ... < i_m} x_{i_1} x_{i_2} \cdots x_{i_m} \cdot \sum_{\tau = (\tau_1 ... \tau_m)} \prod_{t=1}^{m} \left((-1)^{|\tau_t|} (|\tau_t|)! \cdot w_{\tau_t}(i_t)\right) \cdot \mathcal{S}\left(w\left[[a+b] \setminus \cup_t \tau_t\right]\right)$$

Here the inner summation is on all ordered set systems τ such that the terms τ_t are disjoint subsets of $[a+b]$. The terms may also be empty.

Let us attempt to simplify the double summation we obtained. First, we may disregard the constant term $(a! \cdot b!)^{-1}$. Next, observe that, as before, all symmetric functions of the form $\mathcal{S}(w[T])$ vanish, unless T is empty, in which case they equal 1. Therefore, we may consider the double summation

$$\sum_{i_1 < i_2 < ... < i_m} x_{i_1} x_{i_2} \cdots x_{i_m} \cdot \sum_{\tau = (\tau_1 ... \tau_m)} \prod_{t=1}^{m} \left((-1)^{|\tau_t|} (|\tau_t|)! \cdot w_{\tau_t}(i_t)\right)$$

Here the inner summation is on all ordered partitions τ of $[a+b]$. The terms τ_t may also be empty. Changing the order of summation, and ignoring the constant term $(-1)^{a+b}$, we get

$$\sum_{\tau = (\tau_1 ... \tau_m)} \prod_{t=1}^{m} (|\tau_t|)! \cdot \sum_{i_1 < i_2 < ... < i_m} \prod_{t=1}^{m} (x \cdot w_{\tau_t})(i_t) = \sum_{\tau = (\tau_1 ... \tau_m)} \left(\prod_{t=1}^{m} (|\tau_t|)!\right) \cdot \mathcal{F}\left(x w_{\tau_1}, x w_{\tau_2}, ..., x w_{\tau_m}\right)$$

Consider the last expression. Let us use some more notation. For an ordered partition $\tau = (\tau_1 ... \tau_m)$, let $n = n(\tau)$ be the number of empty terms. Let $\{\tau_1 ... \tau_m\}$ denote the unordered version of this partition, where the first $n(\tau)$ terms are taken, by agreement, to be the empty ones. Then we can rewrite this expression as

$$\sum_{\tau=\{\tau_1 ... \tau_m\}} \left(\prod_{t=1}^{m} (|\tau_t|)! \right) \cdot \mathcal{H}\left(x^{(n)}, x w_{\tau_{n+1}}, ..., x w_{\tau_m} \right)$$

Now, clearly not all the terms in the partition are empty and, therefore, $n(\tau) < m$ for all τ, completing the proof of our last claim, of the lemma, and of the theorem. ∎

4 Proof of Theorem 1.3

Let $p = 2$. We will show there is an absolute constant $\alpha > 0$ such that for any polynomial g of degree at most 3 in N variables holds

$$\langle S_4, g \rangle < \exp\{-\alpha N\}$$

A first step is to observe that there is a relation between the inner product of two functions and the average inner product of their derivatives.

Lemma 4.1: *For any two functions f and g holds*

$$\langle f, g \rangle^4 \leq \mathbb{E}_y \langle f_y, g_y \rangle^2$$

Proof: This is an immediate corollary of a lemma in [7], but we give the elementary proof for completeness. By the Cauchy-Schwarz inequality,

$$\mathbb{E}_y \langle f_y, g_y \rangle^2 \geq \mathbb{E}_y^2 \langle f_y, g_y \rangle = \mathbb{E}_{x,y}^2 (-1)^{f(x)+f(x+y)+g(x)+g(x+y)} = \mathbb{E}^4 (-1)^{f(x)+g(x)} = \langle f, g \rangle^4$$

∎

Corollary 4.2:

$$\langle f, g \rangle^8 \leq \mathbb{E}_{y,z} \langle f_{y,z}, g_{y,z} \rangle^2$$

We will show that for any polynomial g of degree at most 3 holds $\mathbb{E}_{y,z} \left\langle (S_4)_{y,z}, g_{y,z} \right\rangle^2 \leq \exp\{-\alpha N\}$. First, here is a brief overview of the argument.

The point is that taking second derivatives makes life easier, since a second derivative of g is a linear function, and a second derivative of S_4 is a quadratic. We therefore need to show that for the large majority of directions y, z, the quadratic function $(S_4)_{y,z}$ has a small inner product with the linear function $(-1)^{g_{y,z}}$. In this we will be helped by a theorem of Dixon giving a structural description of quadratic polynomials, which, in particular, characterizes the Fourier transform of functions of the type $(-1)^Q$, where Q is a quadratic. In fact, setting

$Q = (S_4)_{y,z}$ we will see that for many of the directions y, z the Fourier coefficients of $(-1)^Q$ will be exponentially small. For the remaining directions, these Fourier coefficients will be supported on an explicit easy to describe 3-dimensional affine subspace depending on y, z. We will then argue that for any fixed polynomial g of lower degree, the support of the character $(-1)^{g_{y,z}}$ lies in this affine subspace with exponentially small probability over y, z.

We proceed with computing the second derivative $Q = (S_4)_{y,z}$.

4.1 Second derivatives of S_4

Write $Q(x) = \sum_{i<j} q_{i,j} x(i)x(j) + \sum_i \ell_i x(i) + c$.

By Proposition 2.2 or by Example 2.4.

$$q_{i,j} = \mathcal{S}(y,z) - \langle y, \mathbf{1} \rangle \cdot \Big(z(i) + z(j) \Big) + \langle z, \mathbf{1} \rangle \cdot \Big(y(i) + y(j) \Big) + \Big(y(i)z(j) + y(j)z(i) \Big)$$

At this point we invoke (a corollary of) a theorem of Dixon [6]:

Theorem 4.3: *Let $Q(x) = \sum_{i<j} q_{i,j} x(i)x(j) + \sum_i \ell_i x(i) + c$ be a quadratic polynomial over $\mathbb{F}_2$. Consider the symmetric matrix with zeros on the diagonal and off-diagonal entries given by $S_{i,j} = S_{j,i} = q_{i,j}$. Let the rank of $B = 2h$ (it is always even). Then the function $(-1)^Q$ has 2^{2h} non-zero Fourier coefficients of absolute value 2^{-h}. Moreover, all these coefficients lie in an $2h$-dimensional affine subspace of $\mathbb{F}_2^n$.*

Consider the matrix B in our case. Some notation: let J be the matrix with 0 on the diagonal and 1 off the diagonal. Let $u \otimes v$ denote the outer product uv^t. Then,

$$B = \mathcal{S}(y,z) \cdot J + \langle y, \mathbf{1} \rangle \cdot \Big(z \otimes \mathbf{1} + \mathbf{1} \otimes z \Big) + \langle z, \mathbf{1} \rangle \cdot \Big(y \otimes \mathbf{1} + \mathbf{1} \otimes y \Big) + \Big(y \otimes z + z \otimes y \Big)$$

Since the rank of J is at least $N - 1$ and the rank of the remaining matrices is at most 2, the matrix B is almost of full rank if $\mathcal{S}(y,z) = 1$. In this case, by Theorem 4.3, the Fourier coefficients of $(-1)^Q$ are exponentially small.

We therefore may assume $\mathcal{S}(y,z) = 0$. In this case the quadratic part of Q may be written as

$$\sum_{i<j} q_{i,j} x(i)x(j) = \langle y, \mathbf{1} \rangle \cdot \langle x, \mathbf{1} \rangle \langle x, z \rangle + \langle z, \mathbf{1} \rangle \cdot \langle x, \mathbf{1} \rangle \langle x, y \rangle + \Big(\langle x, y \rangle \langle x, z \rangle + \langle x, yz \rangle \Big)$$

Recall that yz denotes the pointwise product of vectors y and z.

This implies the non-zero Fourier coefficients of $\sum_{i<j} q_{i,j} x(i)x(j)$ lie in a 3-dimensional affine subspace of $\mathbb{F}_2^n$. The linear part of this subspace is spanned by the vectors $y, z, \mathbf{1}$ and it is shifted by a vector yz.

Next, consider the linear part $\sum_i \ell(i)x(i)$ of Q. By Proposition 2.2,

$$\ell(i) = \mathcal{H}^{\{i\}} \Big(y^{(2)}, z \Big) + \mathcal{H}^{\{i\}} \Big(y, z^{(2)} \Big) =$$

$$\sum_{j<k<l\neq i} \Big(y(k)y(l)z(j)+y(j)y(l)z(k)+y(j)y(k)z(l)\Big)+\Big(y(j)z(k)z(l)+y(k)z(j)z(l)+y(l)z(j)z(k)\Big)$$

This can be directly verified to be equal to

$$\Big(\mathcal{S}(y,z)+\mathcal{S}(z,z)+\langle z,\mathbf{1}\rangle\Big)\cdot y(i)+\Big(\mathcal{S}(y,z)+\mathcal{S}(y,y)+\langle y,\mathbf{1}\rangle\Big)\cdot z(i)+$$

$$\Big(\mathcal{S}(y,y)\cdot\langle z,\mathbf{1}\rangle+\mathcal{S}(z,z)\cdot\langle y,\mathbf{1}\rangle+\langle y,z\rangle\cdot\langle y+z,\mathbf{1}\rangle\Big)$$

By assumption, $\mathcal{S}(y,z)=\langle y,\mathbf{1}\rangle\cdot\langle z,\mathbf{1}\rangle+\langle y,z\rangle=0$. Note that this also implies $\langle y,z\rangle\cdot\langle y+z,\mathbf{1}\rangle=0$, implying

$$\ell(i)=\Big(\mathcal{S}(z,z)+\langle z,\mathbf{1}\rangle\Big)\cdot y(i)+\Big((S(y,y)+\langle y,\mathbf{1}\rangle\Big)\cdot z(i)+\Big(\mathcal{S}(y,y)\cdot\langle z,\mathbf{1}\rangle+\mathcal{S}(z,z)\cdot\langle y,\mathbf{1}\rangle\Big)$$

Consequently, the linear part of Q may be written as

$$\sum_i \ell(i)x(i)=$$

$$\Big(\mathcal{S}(z,z)+\langle z,\mathbf{1}\rangle\Big)\cdot\langle x,y\rangle+\Big((S(y,y)+\langle y,\mathbf{1}\rangle\Big)\cdot\langle x,z\rangle+\Big(\mathcal{S}(y,y)\cdot\langle z,\mathbf{1}\rangle+\mathcal{S}(z,z)\cdot\langle y,\mathbf{1}\rangle\Big)\cdot\langle x,\mathbf{1}\rangle$$

This means that the non-zero Fourier coefficients of the polynomial $Q=\sum_{i<j}q_{i,j}x(i)x(j)+\sum_i \ell(i)x(i)+c$ lie in the affine subspace $AF_{y,z}=yz+\mathrm{Span}\,(y,z,\mathbf{1})$.

4.2 Second derivatives of a fixed polynomial of degree 3

Let

$$g(x)=\sum_{i<j<k} a_{i,j,k}x(i)x(j)x(k)$$

be a polynomial of degree 3. For directions $y,z\in\mathbb{F}^N$, consider the second derivative $g_{y,z}=\sum_i v_{y,z}(i)x(i)+c_{y,z}$. We need to show that the probability of the vector $v_{y,z}$ falling in the affine space $AF_{y,z}=yz+\mathrm{Span}\,(y,z,\mathbf{1})$ is exponentially small.

First, some notation. For $1\le i\le N$, let G_i be a symmetric $N\times N$ matrix over $\mathbb{F}$ with $(G_i)_{j,k}=(G_i)_{k,j}=a_{i,j,k}$ for all $j\neq k$. (Here we think about $\{i,j,k\}$ as an unordered subset of $[N]$.) The diagonal entries of G_i are set to 0. For future use note the important property $(G_i)_{j,k}=(G_j)_{i,k}=(G_k)_{i,j}$.

These matrices are relevant because they describe the vector $v_{y,z}$.

Lemma 4.4:

- $$v_{y,z}(i)=coef_{x(i)}\left(g_{y,z}(x)\right)=\langle y,G_iz\rangle$$
- *An alternative representation of $v_{y,z}$ will be more convenient for us. For $z\in\mathbb{F}^N$, let $G(z)=\sum_{i=1}^N z(i)G_i$. Then*
$$v_{y,z}=G(z)\cdot y$$

Proof: For the first claim of the lemma, by linearity of the derivative, it suffices to consider the monomial $g(x) = x(i)x(j)x(k)$. This case can be easily verified directly.

For the second claim, note that

$$(G(z) \cdot y)(l) = \sum_{k=1}^{N} (G(z))_{k,l}\, y(k) = \sum_{k=1}^{N} y(k) \cdot \sum_{i=1}^{N} z(i)\,(G_i)_{k,l} = \sum_{k=1}^{N} y(k) \cdot \sum_{i=1}^{N} (G_l)_{k,i}\, z(i) = \langle y, G_l z\rangle$$

∎

Consider the event $\{v_{y,z} \in AF_{y,z}\}$. This means $v_{y,z} = yz + u_{y,z}$, for some vector $u_{y,z} \in \mathrm{Span}(y, z, \mathbf{1})$. There are only 8 possible choices for $u_{y,z}$. For convenience, let us assume, without loss of generality (as can be easily seen from the proof), that $u_{y,z} = y + z + \mathbf{1}$ is the most popular one. By the lemma, the event $\{v_{y,z} = yz + u_{y,z}\}$ is the same as $\{G(z) \cdot y = yz + u_{y,z}\}$. To simplify things some more, let $A_i = G_i + e_i \otimes e_i$, $i = 1...N$. That is, $A_i = G_i$ but for $(A_i)_{i,i} = 1$. Let $A(z) = \sum_{i=1}^{N} z(i)A_i$. Note that $A(z) \cdot y = G(z) \cdot y + yz$. Hence $\{G(z) \cdot y = yz + u_{y,z}\}$ is the same as $\{A(z) \cdot y = u_{y,z} = y + z + \mathbf{1}\}$

We conclude the proof by a technical claim.

Proposition 4.5: *Let $\{A_i\}$, $i = 1...N$ be a family of symmetric $N \times N$ matrices over $\mathbb{F}$ with $A_i(k,k) = \delta_{ik}$. Then, for y, z uniformly at random and independently from $\mathbb{F}^N$,*

$$Pr_{y,z}\Big\{ (A(z)) \cdot y = y + z + \mathbf{1} \Big\} \le \left(\frac{3}{4}\right)^N$$

The proof of the proposition is based on the claim that the rank of a matrix $A(z)$ is typically large.

Lemma 4.6: *Let matrices $\{A_i\}$ be as in the proposition. Let C be any fixed symmetric $N \times N$ matrix. Then*

$$Pr_z\Big\{ rank(A(z) + C) \le k - 1 \Big\} \le \frac{1}{2^N} \cdot \sum_{i=0}^{k-1} \binom{N}{i}.$$

Proof: Consider a family of $\binom{N}{k}$ polynomials f_I on $\mathbb{F}^N$. These polynomials are indexed by k-subsets of $[N]$. For a k-subset I, let $f_I(z)$ be the determinant of the $I \times I$ minor of $A(z) + C$. Clearly, rank of $A(z) + C$ is smaller than k if and only if z is a joint zero of $\{f_I\}$.

We now claim that the coefficient of $\prod_{i \in I} z_i$ in $f_I(z)$ is 1. If this is true, $\deg(f_I - \prod_{i \in I} z_i) \le k - 1$, and the claim of the lemma will follow from Lemma 4.6.

Let $B(z) = A(z) + C$. Since we are working in characteristic two, the symmetry of $B(z)$ implies that

$$\det B(z) = \sum_{\sigma \in S_N:\ \sigma = \sigma^{-1}} \prod_{i=1}^{N} B_{i\sigma(i)}(z) =$$

$$\sum_{\sigma \in S_N:\ \sigma = \sigma^{-1}} \prod_{\{i: \sigma(i) = i\}} (z_i + C_{i,i}) \cdot \prod_{\{i: i < \sigma(i)\}} B_{i\sigma(i)}(z) = \prod_{i \in I}^{n} z_i + \text{lower order terms.}$$

In the second equality we use the identity $B^2_{i\sigma(i)}(z) = B_{i\sigma(i)}(z)$ in $\mathbb{F}$. ▮

Let I denote the identity $N \times N$ matrix.

Let $p(z) = Pr_y\Big\{A(z) \cdot y = y + z + \mathbf{1}\Big\}$. Clearly $p(z) \leq 2^{-\text{rank}(A(z)+I)}$. By Lemma 4.6,

$$Pr_{y,z}\Big\{\ (A(z)) \cdot y = y + z + \mathbf{1}\Big\} = \mathbb{E}_z p_z \leq \mathbb{E}_z 2^{-\text{rank}(A(z)+I)} \leq \frac{1}{2^N}\sum_{k=0}^{N}\binom{N}{k}2^{-k} = \left(\frac{3}{4}\right)^N$$

This concludes the proof of the proposition, and of Theorem 1.3.

5 Acknowledgements

We are grateful to Ben Green and Terence Tao for sending us their paper. We would also like to thank them and Emanuele Viola for bringing the argument of Alon and Beigel to our attention

References

[1] N. Alon, R. Beigel *Lower Bounds for Approximations by Low Degree Polynomials over Z_m*, SCT: Annual Conference on Structure in Complexity Theory, 2001.

[2] A. Bogdanov, E. Viola *Pseudorandom bits for polynomials via the Gowers norm*, FOCS'07.

[3] W. T. Gowers, *A new proof of Szemeredi's theorem*, GAFA Vol. 11(2001), pp. 465-588.

[4] B. Green, T. Tao, *An inverse theorem for the Gowers U^3 norm*, Proc. Edinburgh Math. Soc., to appear.

[5] B. Green, T.Tao, *The distribution of polynomials over finite fields, with applications to the Gowers norms*, preprint, 2007.

[6] J. MacWilliams and N. J. A. Sloane, **The Theory of Error Correcting Codes**, Amsterdam, North-Holland, 1977.

[7] A. Samorodnitsky, *Low degree tests at large distances*, STOC '07.

[8] T. Tao, *Structure and randomness in combinatorics*, FOCS '07.

[9] E. Viola, A. Wigderson, *Norms, XOR lemmas, and lower bounds for $GF(2)$ polynomials and multiparty protocols*, CCC '07.