Inverse Conjecture for the Gowers Norm is False

Shachar Lovett  
Faculty of Mathematics and Computer Science  
The Weizmann Institute of Science  
Rehovot 76100, Israel  
shachar.lovett@weizmann.ac.il

Roy Meshulam  
Department of Mathematics  
The Technion  
Haifa 32000, Israel  
meshulam@math.technion.ac.il

Alex Samorodnitsky  
School of Engineering and Computer Science  
The Hebrew University of Jerusalem  
Jerusalem 91904, Israel  
salex@cs.huji.ac.il

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 \( f : \mathbb{F}_p^N \to \mathbb{F}_p \) is non-negligible, that is larger than a constant independent of \( N \), then \( f \) has a non-trivial correlation with 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 with any polynomial of degree 3 is exponentially small. Essentially the same result, with different bounds for correlation, was independently obtained by Green and Tao [8]. 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
This paper belongs to the framework of the "local-to-global" research theme in computer science. The high-level aim of this research direction is to study the global correlation of a general combinatorial object, such as a boolean function or a graph, with a specific subclass \( S \) of structured objects of the same type, by means of the percentage of some local tests that the object passes.

These local tests usually have the following convenient properties: any structured object from \( S \) passes all the local tests. On the other hand, if an object passes all the local tests then it lies in \( S \).

The challenging goal is to infer non-trivial correlation with some object in \( S \) based on some non-trivial fraction of the tests succeeding. We observe that another natural framework for this type of questions is that of Property Testing [5], and, in fact, if the described goal is achieved, this makes the property \( S \) locally testable in the appropriate region of correlation (or distance) with objects from \( S \).

In this paper the objects we consider are multivariate functions from a finite field to itself, and the structured objects are low-degree multivariate polynomials.

A variant of this question was considered in the proof of the PCP theorem. In the PCP setting, the field size is larger than the degree and typical tests check that the restriction of the function to a line or a plane is a low-degree univariate or bivariate polynomial.

Here we consider the case of a small field (smaller than the degree), which does not allow us to use these local tests. Instead, we consider another local condition that a degree \( d-1 \) polynomial must satisfy - that its properly defined \( d \)-th directional derivative vanishes for any set of \( d \) directions. This leads to a local condition on affine cubes of dimension \( d \), and has been studied in several works on Reed-Muller codeword testing [2, 10, 12]. This problem, in diverse algebraic settings, was also independently considered in additive combinatorics [7] following Gowers' proof of Szemerédi's theorem [6] and in ergodic theory [9].

This line of research led to the conjecture [7, 12, 13], that if the behaviour of a typical \( d \)-th directional derivative of a function \( f \) differs non-trivially from that of a random (i.e., unstructured) function, than \( f \) has a non-trivial correlation with a polynomial of degree \( \geq (d-1) \). This conjecture is known to hold for \( d = 2 \) [4, 3], and in the case where the derivative vanishes with probability very close to 1, depending on the field size and the degree of the polynomial, [2, 10]. The general case for \( d = 3 \) has been established independently.
in [7] for prime fields of odd order and in [12] for the binary field.

The main result of this paper is that the conjecture is false for \( d = 4 \).

We now pass to a more detailed description of the setting and of our results.

Let \( p \) be a prime number. Let \( \mathbb{F} = \mathbb{F}_p \) be the finite field with \( p \) elements. In the following, the field \( \mathbb{F} \) will be fixed and we will consider multivariate functions \( f : \mathbb{F}^N \to \mathbb{F} \) for various values of \( N \). A special subclass of multivariate functions is that of low-degree polynomials. Let \( d \) be an integer number, and let \( S \) denote the vector space of all \( N \)-variate degree-(\( d - 1 \)) polynomials over \( F \). We will think about \( d \) as fixed, depending on the field but not on the number of variables \( N \). The main question we consider is how to decide whether a given \( N \)-variate function \( f \) has a non-trivial correlation with a low-degree polynomial from \( S \) by performing some local tests on \( y \) if \( f \) is non-trivial.

We will test \( f \) by computing a 'random \( d \)-th directional derivative of \( f \). To be more specific, for 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 \( k \)-th derivative in these directions to be
\[
f_{y_1...y_k} = (f_{y_1...y_{k-1}})_{y_k}
\]
It is easy to see that this definition does not depend on the ordering of \( y_1...y_k \).

A very useful technical tool to measure the behavior of a random \( k \)-th directional derivative of \( f \) is the Gowers uniformity norm introduced by Gowers in [6]. Let \( \xi = e^{2\pi i} \) be the primitive \( p \)-th root of unity. Denote by \( e(x) \) the exponential function taking \( x \in \mathbb{F} \) to \( \xi^x \in \mathbb{C} \).

The \( k \)-th Gowers "norm" \( \|f\|_{U^k} \) of \( f \) is
\[
\|f\|_{U^k} = \left( \mathbb{E}_{x \in \mathbb{F}^n} [e(f_{y_1...y_k}(x))] \right)^{1/2^k}
\]
More accurately, as shown in [6], this is indeed a norm of the associated complex-valued function \( e(f) \) for \( k \geq 2 \).

It is easy to see that \( \|f\|_{U^0} = 1 \) if \( f \) is a polynomial of degree at most \( d - 1 \). This is just another way of saying that all order-\( d \) iterative derivatives of \( f \) are zero if and only if \( f \) is a polynomial of degree at most \( d - 1 \). To present a strengthening of this result we define a correlation of two functions \( f \) and \( g \) to be
\[
|\langle f, g \rangle| = \left| \mathbb{E}_x e(f(x) - g(x)) \right|
\]
Then we have [7]: If \( f \) has a non-trivial correlation with a low-degree polynomial, that is \( |\langle f, g \rangle| \geq \epsilon \) for a polynomial \( g \) of degree at most \( d - 1 \), then \( \|f\|_{U^d} \geq \epsilon \).

In the other direction, we have the inverse conjecture [7, 12, 13], that if \( d \)-th Gowers norm of \( f \) is non-trivial, then \( f \) is has a non-trivial correlation with a degree-(\( d - 1 \)) polynomial. More formally,

The "Inverse Conjecture for the Gowers Norm" Let \( \epsilon > 0 \) and let \( f : \mathbb{F}^n \to \mathbb{F} \) with \( \|f\|_{U^d} \geq \epsilon \). Then there exists \( \epsilon' = \epsilon'(\mathbb{F}, d, \epsilon) > 0 \) and a degree-(\( d - 1 \)) polynomial \( g \) such that \( |\langle f, g \rangle| \geq \epsilon' \).

In this paper we show this conjecture, which we will refer to as "ICGN", to be false by providing a counterexample. This counterexample comes from a special class of functions: symmetric polynomials. The elementary symmetric polynomial of degree \( n \) in \( N \) variables is defined as
\[
S_{n,N}(x) = \sum_{S \subseteq [N], |S| = n} \prod_{i \in S} x_i
\]
In the following we will usually omit the subscript \( N \).

We prove two types of claims about symmetric polynomials.

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

**Theorem 1.**
- For any prime \( p \), there is an absolute positive constant \( \epsilon = \epsilon(p) \) such that for the symmetric polynomial \( S_{2p} \) over \( \mathbb{F}_p \) holds
  \[
  \|S_{2p}\|_{U^{p+2}} > \epsilon,
  \]
- An important special case is that of the binary field: \( p = 2 \)
  \[
  \|S_{2}\|_{U^4} > \epsilon
  \]  
  (1)
- An easy generalization: for any \( n \geq 2p \),
  \[
  \|S_n\|_{U^{n-p+2}} > \epsilon
  \]  
  (2)

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

**Theorem 2.** Let \( p = 2 \). There is an absolute constant \( \alpha > 0 \) such that for any polynomial \( g \) of degree \( 3 \) holds
\[
|\langle S_4, g \rangle| < \exp(-\alpha N)
\]  
(3)
We conjecture the claim of the theorem to be true for any prime number \( p \), replacing \( 3 \) with \( p + 1 \), \( 4 \) with \( 2p \), and the constant \( \alpha \) with an appropriate constant \( \alpha' \) depending on \( p \).

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

For the case of general prime \( p \), (2) and the monotonicity of the Gowers norm ((7)) directly imply that:
\[
\|S_{2p}\|_{U^{p+2}} > \epsilon
\]
We observe that a minor adaptation of the Alon-Beigel argument, as also used in [8] for the case of \( S_4 \), shows that \( S_{2p} \) cannot be approximated by lower degree polynomials.

**Lemma 3.** Let \( p > 2 \) be a prime. For any polynomial \( g \) of degree \( p^2 - 1 \) holds
\[
|\langle S_{p^2}, g \rangle| = o_N(1)
\]  
(4)
Combining these two results shows the ICGN to be false for any prime \( p \) and \( d = p^2 \).

### 1.1 Related work

Our results have a large overlap with a recent work of Green and Tao [8].

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, in its full generality, is shown to be false. In particular, the symmetric polynomial \( S_4 \) is shown to be a counterexample for \( p = 2 \) and \( d = 4 \).

The proof of non-approximability of \( S_4 \) by lower-degree polynomials in [8] 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 than these obtained in this paper for non-inapproximability of $S_b$, in that it shows $(S_b, g) < \log^{-c}(N)$ for an absolute constant $c > 0$ and any degree-3 polynomial $g$.

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 $p$.

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 is proved in Section 3. Theorem 2 is proved in Section 4. Lemma 3 is proved in Section 5.

2. PRELIMINARY NOTIONS AND CLAIMS

2.1 Properties of some multilinear polynomials

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 $r_1, \ldots, r_n$ will be denoted by $M(r_1 \ldots r_n)$. Sometimes there will be repeated rows. In such a case we consider a partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ of $[n]$, that is $\lambda_i$ are (possibly empty) subsets of $[n]$, whose disjoint union is $[n]$. We denote by $M_{\lambda}(r_1 \ldots r_k)$ the matrix whose rows in positions indexed by elements of $\lambda_i$ equal $r_i$.

Note that the partition $\lambda$ is ordered, in that the ordering of its terms $\lambda_i$ is important. We use the notation $\{\lambda_1, \ldots, \lambda_k\}$ for an unordered partition.

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

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

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

Let $\lambda = (\lambda_1, \ldots, \lambda_k)$ be a partition of $[n]$, and let $\ell_i = |\lambda_i|$. Clearly $S(M_{\lambda})$ depends only on the cardinalities $\ell_i$ of $\lambda_i$. This leads to the notation $M \left[ r_1^{(\ell_1)} \ldots r_k^{(\ell_k)} \right]$ which denotes the matrix in which the row $r_1$ appears $\ell_1$ times, followed by the $\ell_2$ appearances of the row $r_2$ and so on. In this notation, therefore

$$S(M_{\lambda})(r_1 \ldots r_k) = S \left( M \left[ r_1^{(\ell_1)} \ldots r_k^{(\ell_k)} \right] \right)$$

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

$$\mathcal{F}(M)[r_1 \ldots r_n] = \sum_{C \subseteq [n], |C| = \lambda} \prod_{i=1}^{n} r_i(j_i)$$

Here $r_i(j)$ denote the $j$-th coordinate of the vector $r$. To connect the two notions, observe that

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

where $\sigma$ 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 $\lambda = (\lambda_1, \ldots, \lambda_k)$ be an ordered partition of $[n]$ with $k$ terms. For any such partition $\theta = (\theta_1, \ldots, \theta_k)$ of $[n]$ write $\theta \sim \lambda$ if $|\theta_1| = |\lambda_1|, \ldots, |\theta_k| = |\lambda_k|$. We define

$$\mathcal{H}(M_{\lambda}(r_1 \ldots r_k)) = \sum_{C \subseteq [n], |C| = \lambda} \sum_{\theta \sim \lambda} \prod_{i=1}^{k} r_i(j_i)$$

An alternative view of the functions $S, \mathcal{F}$ 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 $\rho$ monotone on a subset $\{i_1 < i_2 < \ldots < i_\ell\}$ of $[n]$ if $\rho(i_1) < \rho(i_2) < \ldots < \rho(i_\ell)$. A path is (fully) monotone if it is monotone on $[n]$. Then, for a partition $\lambda = (\lambda_1, \ldots, \lambda_k)$ of $[n]$ and an $n \times N$ matrix $M = M_{\lambda}$,

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

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

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

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

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

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

$$S(M) = \left( \prod_{i=1}^{k} |\lambda_i|! \right) \cdot \mathcal{H}(M) \quad (5)$$

Note that this is an identity in $\mathbb{F}$. In particular, if one of the terms $\lambda_i$ has cardinality at least $p$ then $S(M) = 0$ and (5) provides no information.

To simplify the notation we will usually write $S(r_1 \ldots r_n)$ for $S(M_{\lambda}(r_1 \ldots r_k))$, $\mathcal{F}_{\lambda}(r_1 \ldots r_k)$ for $\mathcal{F}(M_{\lambda}(r_1 \ldots r_k))$ 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 $M[x_1 \ldots x_n]$, 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 $S_{n}(x_1 \ldots x_k)$ 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 $y_i$. 549
The basic observation here is the following lemma which is straightforward from the definition of directional derivative.

**Lemma 4.** Let a polynomial $P(x)$ in $N$ variables be given by

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

Then

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

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 5.** Let $k \leq n$, then

$$(S_n)(y_1 \ldots y_k)(x) = \sum_{m=0}^{n-k} \mathcal{H}(x^{(m)}, y_1^{(\ell_1)} \ldots y_k^{(\ell_k)})$$

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

$$\sum_{m=0}^{n-k} \mathcal{H}(x, y_1 \ldots y_k) = \sum_{m=0}^{n-k} \mathcal{H}(x^{(m)}, y_1^{(\ell_1)} \ldots y_k^{(\ell_k)})$$

**Proof.** Iterating Lemma 4, 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 $x^1 \cdots x^m$ in $(S_n)(y_1 \ldots y_k)$.

**Corollary 6.**

1. $a = \mathcal{H}^{(j_1 \ldots j_m)}(y_1^{(\ell_1)} \ldots y_k^{(\ell_k)})$
2. If $k + m + p > n + 1$ then

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

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 5. The second claim follows from the first claim, from (5), 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 1.** 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 $P = (S_4)^{x, z}$ Then $P$ is a quadratic polynomial and for $1 \leq i < j \leq N$

$$\text{coeff}_{i,j}(P) = \sum_{k \neq i, k \neq j} y(k)z(l) = \mathcal{S}(i, j)(y, z)$$

Continuing with the same example, note that it convenient to express the symmetric function $S(y, z)$ via the complete symmetric function $S(y, z)$ minus the symmetric function $S(y, z)$ minus some of the terms, as follows: $S(i, j)(y, z) = S(y, z) - (y + z)(j) + (y)(i + j) + (y)(i + j)$.

Note the "inclusion-exclusion" structure in the two expressions above. (To make it even clearer we use "n" and "r" 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 $y_1 \ldots y_k$.

### 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 $y_\tau$ be the vector whose coordinate $i$-th coordinate of the vectors $y_1 \cdots y_m$.

For the value of the function $\mathcal{S}$ on a matrix with $|\tau|$ rows $y_\tau$, $i \in \tau$. Let $S(y_\tau)$ be the polynomial $S(y_\tau)$.

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

**Lemma 7.**

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

From now on we assume $r_0$ to be the all-1 vector, and $S(r[0])$ to equal 1.

**Proof.** The proof is by induction on $n$. For $n = 1$ both sides equal $\sum_{i=1}^{n} r_i(1) - r_i(k)$. For $n > 1$, observe that

$$S^{(k)}(r_1 \ldots r_n) = S(r_1 \ldots r_n) - \sum_{i=1}^{n} r_i(k) \cdot S^{(k)}(r[|n| \setminus \{i\}])$$

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 $S(r_1\ldots r_n)$ via inner products $\langle r_T \rangle$.

**Proposition 8.**

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

In this summation $\lambda = (\lambda_1\ldots \lambda_m)$ runs over all unordered partitions of $[n]$ with non-empty $\lambda_t$.

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

\[
S(r_1\ldots r_n) = \sum_{k=1}^N r_n(k) \cdot S^{(k)}(r_1\ldots r_{n-1})
\]

Using Lemma 7 and the induction hypothesis,

\[
S(r_1\ldots r_n) = \sum_{k=1}^N r_n(k) \cdot \sum_{\tau \subseteq [n-1]} (-1)^{|\tau|} \cdot \langle r_{\tau} \rangle \cdot S\left( [n-1] \setminus \tau \right)
\]

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

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

\[
S(r_1\ldots r_n) = \sum_{\tau \subseteq [n-1]} (-1)^{|\tau|} \cdot \langle r_{\tau} \rangle \cdot S\left( [n-1] \setminus \tau \right)
\]

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

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

\[
S(r_1\ldots r_n) = \sum_{\tau \subseteq [n-1]} (-1)^{|\tau|} \cdot \langle r_{\tau} \rangle \cdot S\left( [n-1] \setminus \tau \right)
\]

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

\[
S(r_1\ldots r_n) = \sum_{\tau \subseteq [n-1]} \prod_{t=1}^m \left((-1)^{|\lambda_t|-1}(|\lambda_t| - 1)! \cdot \langle r_{\lambda_t} \rangle \right)
\]

completing the proof of the proposition. □

The second claim expresses the incomplete symmetric function $S^{(j_1\ldots j_k)}(r_1\ldots r_n)$ as a polynomial in the missing coordinates $j_1\ldots j_k$ of the vectors $r_i$ and in complete symmetric functions applied to sub-matrices of $M(r_1\ldots r_n)$. Note that Lemma 7 is a special case $k = 1$ of this claim.

**Proposition 9.**

\[
S^{(j_1\ldots j_k)}(r_1\ldots r_n) = \sum_{\tau = (r_1\ldots r_k)} \prod_{t=1}^k \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t) \right) \cdot S\left( [n] \setminus \cup \tau_t \right)
\]

Here the summation is on all ordered set systems $\tau$ such that the terms $\tau_t$ 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 7.

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

\[
\sum_{\tau = (r_1\ldots r_k)} \prod_{t=1}^k \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t) \right) \cdot S\left( [1] \setminus \cup \tau_t \right)
\]

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

For $k, n > 1$, we have

\[
S^{(j_1\ldots j_k)}(r_1\ldots r_n) = S^{(j_1\ldots j_{k-1})}(r_1\ldots r_{n-1}) - \sum_{i=1}^n r_i(j_k) \cdot S^{(j_1\ldots j_{k-1})}(r_{[n]} \setminus \{i\})
\]

By the induction hypothesis, this is

\[
\sum_{\tau = (\theta_1\ldots \theta_{k-1})} \prod_{t=1}^{k-1} \left((-1)^{|\theta_t|}(|\theta_t|)! \cdot r_{\theta_t}(j_t) \right) \cdot S\left( [n] \setminus \cup \theta_t \right)
\]

\[
- \sum_{\tau = (\theta_1\ldots \theta_{k-1})} \prod_{t=1}^{k-1} \left((-1)^{|\theta_t|}(|\theta_t|)! \cdot r_{\theta_t}(j_t) \right) \cdot S\left( [n] \setminus \cup \theta_t \right)
\]

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

Given a set system $\theta = (\theta_1\ldots \theta_{k-1})$ we define a set system $\tau = (\tau_1\ldots \tau_k)$ by setting $\tau_t = \theta_t$, $t = 1\ldots k-1$ and $\tau_k = \emptyset$. Given a set system $\mu^{(i)} = (\mu^{(i)}_1\ldots \mu^{(i)}_k)$ we define a set system $\tau = (\tau_1\ldots \tau_k)$ by setting $\tau_u = \mu^{(i)}_u$, $u = 1\ldots k-1$ and $\tau_k = \emptyset$. 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 $\theta$-system, and each system with non-empty $k$-th term $\tau_k$ is obtained exactly $|\tau_u|$ times, from systems $\mu^{(i)}$ with $i \in \tau_u$. Rearranging the terms and the signs, the last expression is precisely

\[
\sum_{\tau = (r_1\ldots r_k)} \prod_{t=1}^k \left((-1)^{|\tau_t|}(|\tau_t|)! \cdot r_{\tau_t}(j_t) \right) \cdot S\left( [n] \setminus \cup \tau_t \right)
\]

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 10.** Let \( K \) be an absolute constant. Let \( y_{i,j}, i = 1, \ldots, p - 1, j = 1, \ldots, K \), 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

\[
||f||_{U^p} > \left( \frac{|M|}{2^N} \right)^2 =: P_{r^2}(M)
\]

**Proof.** Let \( f_M \equiv 0 \).

Consider a subspace \( V \) of polynomials of degree at most \( p - 1 \) in \( \mathbb{F}[x_1, \ldots, x_1] \) spanned by the polynomials \( \langle x^i, y_{i,j} \rangle \), for all \( i,j \). We will first find a polynomial \( q \in V \) such that \( |\langle f, q \rangle| \geq P_{r^2}(M) \). This, combined with a lemma from \([7]\), will imply the claim of the proposition.

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

\[
\mu(b,c) = Pr\{x : f(x) = c \land \langle x^i, y_{i,j} \rangle = b_{i,j} \text{ for all } i,j\}
\]

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

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

\[
\langle f, g \rangle = \mathbb{E}e(f - g) = \sum_{b,c} \mu(b,c) \cdot e(c - \langle a, b \rangle)
\]

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

\[
\mathbb{E}g \in V | \langle f, g \rangle \geq \frac{1}{|V|} \sum_{b,c} \mu(b,c) \cdot e(c - \langle a, b \rangle) = \frac{1}{|V|} \sum_{b,c} \mu(b,c) \cdot e(c) \cdot e(-\langle a, b \rangle)
\]

\[
= \frac{1}{|V|} \sum_{b,c} \mu(b,c) \cdot e(c) \cdot \mu(0,c_0) = \mu(0,c_0) \cdot e(c_0) = Pr(M) \cdot e(c_0)
\]

This means, there is \( g \in V \) with \( |\langle f, g \rangle| \geq P_{r^2}(M) \). We conclude the proof of the proposition by quoting a lemma from \([7]\), 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 multivariate polynomials. Let \( n \) be fixed integer and let \( N \) be an integer parameter growing to infinity. Let \( r_1, \ldots, r_n \) be \( n \) vectors in \( \mathbb{F}^N \). Let \( \kappa = (k_1, \ldots, k_n) \) be a non-zero sequence of integers \( 0 \leq k_i < p \). For each such sequence define a polynomial \( X_\kappa(x_1, \ldots, x_n) = \sum_{j=1}^N \prod_{i=1}^n r_i^{k_i}(j) \).

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

**Proposition 11.** Let \( U \) be the uniform distribution on \( \mathbb{F}^N \). Let \( P \) be distribution of \( X \) on \( \mathbb{F}^N \). Let \( ||\cdot|| \) denote the statistical \( (1/1) \) 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 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 \( \chi(y) = \xi(y, \omega) \), corresponding to a non-zero vector \( a = (a_\kappa)_\kappa \in \mathbb{F}^K \). (Recall that \( \xi = e^{2\pi i/p} \) is the \( p \)-th primitive root of unity.) Then, normalizing appropriately,

\[
\hat{P}(\chi) = \sum_{y} P(y) \bar{\chi}(y) = \sum_{y} Pr(X = y) \cdot e^{-\sum \langle a_\kappa y_\kappa \rangle} = \mathbb{E}\xi^{-\sum a_\kappa X_\kappa}
\]

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

We have

\[
X_\kappa(r_1, \ldots, r_n) = \sum_\kappa a_\kappa P_\kappa(r_1, \ldots, r_n) = \sum_{j=1}^N \sum_{\kappa=(k_1, \ldots, k_n)} a_\kappa \prod_{i=1}^n r_i^{k_i}(j)
\]

Let \( x_i \) be elements of the field \( F \). Consider an \( n \)-variate polynomial

\[
Q(x_1, \ldots, x_n) = \sum_{\kappa=(k_1, \ldots, k_n)} a_\kappa \prod_{i=1}^n x_i^{k_i}
\]

Since not all of the coefficients \( a_\kappa \) are zero, and since all \( \kappa \) are non-zero sequences, \( Q \) is a multi-variate polynomial of degree at least 1 in \( \mathbb{F}[x_1, \ldots, x_n] \), and therefore attains at least two values with probability bounded away from zero. Now, \( X_\kappa = \sum_{j=1}^N Q(r_1(j), \ldots, r_n(j)) \) is a sum of \( N \) independent copies of \( Q \). Let \( \mu \) denote the distribution of \( Q \) on \( \mathbb{F} \). Then the distribution \( P_\kappa \) of \( X_\kappa \) is \( \mu^p \), the \( N \)-wise convolution of \( \mu \) with itself. Since \( p \) is prime, \( \hat{\mu}(0) = 1 \), and \( \hat{\mu} < 1 \) everywhere else. Therefore, \( \hat{P}_\kappa = (\hat{\mu})^N \) 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 12.** Let \( M \) be the ring of \( \mathbb{F} \)-valued functions on \( \mathbb{F}^N \), that is \( M = \mathbb{F}[x_1, \ldots, x_n]/I \), where \( I \) is the ideal...
$(x_1^n - x, ..., x_N^n - x)$. Let $f_1...f_K$ be polynomials in $M$. Let $S$ be the set of common zeroes of $f_1...f_K$, that is

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

Then $|S| \leq \dim(M/J)$, where $J$ is the ideal generated by $\{f_i\}$, and $\dim(M/J)$ denotes the dimension of $M/J$, 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 $q(v)$ 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. \(\square\)

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

**Lemma 13.** 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

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

(6)

Then,

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

**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 $k - 1$. \(\square\)

3. PROOF OF THEOREM 1

We need to show that

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

for an absolute constant $\epsilon$.

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

Recall (17) that $\|f\|_{U^{p+2}} = \mathbb{E}_{y,z} \|f_{y,z}\|_{U^{p}}$. Since the Gowers’ norms are non-negative, to show that $\|f\|_{U^{p+2}}$ is non-negligible, it suffices to show that $\|f_{y,z}\|_{U^{p}}$ is non-negligible for a non-negligible fraction of directions $y, z \in \mathbb{F}^N$.

Let

$$A = \{(y, z) : \langle y^a, z^b \rangle = 0 \text{ for all } 0 \leq a, b < p, \ a + b > 0\}$$

By Proposition 11, 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 $\epsilon'$.

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

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

In Lemma 14 we will show that $f_{y,z}$ is fixed on $M$. Assuming this, by Proposition 10, we have $\|f_{y,z}\|_{U^{p}} > Pr^2(M)$, and therefore

$$\|f\|_{U^{p+2}} = \mathbb{E}_{y,z} \|f_{y,z}\|_{U^{p}} \geq \Pr\{A\} \cdot \mathbb{E}_{(y,z) \in A} \Pr\{M(y, z)\} \geq \left(\Pr\{A\} \cdot \mathbb{E}_{(y,z) \in A} \Pr\{M(y, z)\}\right)^{2p+1} = \Pr\{x : \langle x^a, z^b \rangle = 0, 0 \leq a, b, i \leq p - 1\} \geq \Omega(p^{-p^2})$$

The last inequality follows from Proposition 11, since random variables $\langle x^a, z^b \rangle$ are asymptotically uniform and independent. It remains to prove the following fact.

**Lemma 14.** Let $x, y, z$ be three vectors in $\mathbb{F}^N$ satisfying $\langle x^a, y^a, z^b \rangle = 0$ for all $0 \leq a, b, i \leq p - 1$. Then

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

**Proof.** By Proposition 5,

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

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

There are two possible cases to consider. The easier case is when $a, b, m < p$. In such a case, by (5), $\mathcal{H}(x^{(m)}, y^{(a)}, z^{(b)})$ is proportional to $S(x^{(m)}, y^{(a)}, z^{(b)})$. By Proposition 8, the symmetric function $S(x^{(m)}, y^{(a)}, z^{(b)})$ is a polynomial in $\langle x^a, y^a, z^b \rangle$, which vanishes when all of these inner products are zero.

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 $\mathcal{H}(x^{(m)}, y^{(a)}, z^{(b)})$ can be written as a linear combination of hybrid functions $\mathcal{H}(x^{(\ell)}, r_1,..., r_{m-\ell})$, where $\ell < m$ and the vectors $r_i$ are of the form $x^a y^a z^7$. Note that this will suffice to prove the lemma, since iterating this step will express $\mathcal{H}(x^{(m)}, y^{(a)}, z^{(b)})$ as a linear combination of symmetric functions in $r_i$, and these functions vanish.

Consider $\mathcal{H}(x^{(m)}, y^{(a)}, z^{(b)})$. 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 6 and Proposition 9,
\[ H \left( (x^{(m)}, y^{(a)}, z^{(b)}) \right) = \]
\[(a! \cdot b!)^{-1} \cdot \sum_{i_1 < i_2 < \ldots < i_m} x_{i_1} x_{i_2} \ldots x_{i_m} S^{(i_1 \ldots i_m)} \left( y^{(a)}, z^{(b)} \right) = \]
\[(a! \cdot b!)^{-1} \cdot \sum_{i_1 < i_2 < \ldots < i_m} x_{i_1} x_{i_2} \ldots x_{i_m} \cdot \sum_{\tau = (\tau_1 \ldots \tau_m)} \prod_{t=1}^m \left( \frac{\tau_t!}{(\tau_t)!} \cdot w_{\tau_t} (i_t) \right) \cdot \]
\[ S \left( w[a \cup b] \cup \bigcup_{t=1}^m \tau_t \right) \]

Here the inner summation is on all ordered set systems \( \tau \) such that the terms \( \tau_t \) 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 \( S(w[T]) \) vanish, unless \( T \) is empty, in which case they equal 1. Therefore, we may consider the double summation
\[
\sum_{i_1 < i_2 < \ldots < i_m} x_{i_1} x_{i_2} \ldots x_{i_m} \cdot \sum_{\tau = (\tau_1 \ldots \tau_m)} \prod_{t=1}^m \left( \frac{\tau_t!}{(\tau_t)!} \cdot w_{\tau_t} (i_t) \right) \]

Here the inner summation is on all ordered partitions \( \tau \) of \([a+b]\). The terms \( \tau_t \) may also be empty. Changing the order of summation, and ignoring the constant term \((-1)^{a+b}\), we get
\[
\sum_{\tau = (\tau_1 \ldots \tau_m)} \prod_{t=1}^m \left( \frac{\tau_t!}{(\tau_t)!} \right) \cdot \sum_{i_1 < i_2 < \ldots < i_m} \prod_{t=1}^m (x \cdot w_{\tau_t} (i_t)) = \]
\[
\sum_{\tau = (\tau_1 \ldots \tau_m)} \prod_{t=1}^m \left( \frac{\tau_t!}{(\tau_t)!} \right) \cdot \mathcal{F}(x_{\tau_1} w_{\tau_1}, x_{\tau_2} w_{\tau_2}, \ldots, x_{\tau_m} w_{\tau_m}) \]

Consider the last expression. Let us use some more notation.

For an ordered partition \( \tau = (\tau_1 \ldots \tau_m) \), let \( n = n(\tau) \) be the number of empty terms. Let \( \{\tau_1 \ldots \tau_m\} \) 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
\[
\sum_{\tau = (\tau_1 \ldots \tau_m)} \left( \prod_{t=1}^m \frac{\tau_t!}{(\tau_t)!} \right) \cdot H \left( (x^{(m)}, x_{\tau_1} w_{\tau_1}, \ldots, x_{\tau_m} w_{\tau_m}) \right) \]

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

4. PROOF OF THEOREM 2

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 15. For any two functions \( f \) and \( g \) holds
\[ \langle f, g \rangle^4 \leq E_y \langle f_y, g_y \rangle^2 \]

Proof. This is an immediate corollary of a lemma in [12], but we give the elementary proof for completeness. By the Cauchy-Schwarz inequality, \( E_y \langle f_y, g_y \rangle^2 \geq E_y^2 \langle f_y, g_y \rangle = E_{y,z} (-1)^{f(x)+f(x+y)+g(x)+g(x+y)} = E^4(-1)^{f(x)+g(x)} = \langle f, g \rangle^4 \]

\[ \Box \]

Corollary 16.
\[ \langle f, g \rangle^8 \leq E_{y,z} \langle f_y, g_y \rangle^2 \]

We will show that for any polynomial \( g \) of degree at most 3 holds \( E_{y,z} \langle S_4 \rangle_{y,z}, g_y, g_z \rangle^2 \leq \exp(-\alpha N) \). 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 function \((-1)^Q y, z \). 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 \((-1)^Q g \) lies in this affine subspace with exponentially small probability over \( y, z \).

Let us observe, with regards to the preceding argument, that going further on and taking third derivatives of \( g \) and \( S_4 \) would not be useful here. Indeed, it is not hard to see that a random third derivative of \( S_4 \) is constant with positive probability independent from \( N \). Therefore an application of the appropriate extension of the Corollary would upper bound the correlation of \( g \) and \( S_4 \) by an absolute constant, which is not what we need.

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 5 or by Example 1, \( q_{i,j} = S(y, z) - (y, 1) \cdot (z+i) + (z, 1) \cdot (y+i)+y) + (y+i+z)+(j+i+z) \).

At this point we invoke (a corollary of) a theorem of Dixon [11]:

Theorem 17. 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 \( F_2 \). 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 \( F_2^6 \).

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 $B = S(y,z) \cdot J + \langle y,1 \rangle \cdot (z \otimes 1 + 1 \otimes z) + \langle z,1 \rangle \cdot (y \otimes 1 + 1 \otimes y) \cdot (y \otimes z + z \otimes y)$.

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 $S(y,z) = 1$. In this case, by Theorem 17, the Fourier coefficients of $(−1)^N$ are exponentially small.

We therefore may assume $S(y,z) = 0$. In this case the quadratic part of $Q$ may be written as $\sum_{i<j} q_{ij} x(i)x(j) = \langle y,1 \rangle \cdot (x,1) \cdot (x,z) + \langle z,1 \rangle \cdot (x,1) \cdot (x,y) + \langle x, y, z \rangle$.

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

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

Next, consider the linear part $\sum_i \ell(i)x(i)$ of $Q$. By Proposition 5,

$$\ell(i) = \mathcal{H}^{(1)}(y^{(2)}, z) + \mathcal{H}^{(1)}(y, z^{(2)}) = \sum_{j<k \neq i} \left( (y(j)y(l)z(j) + y(j)y(l)z(k) + y(j)y(k)z(l)) + (y(j)z(k)z(l) + y(k)z(j)z(l) + y(l)z(j)z(k)) \right)$$

This can be directly verified to be $\left( S(y,z) + S(z,y) + S(z,z) + (z,1) \right) \cdot y(i) + \left( S(y,y) + S(y,z) + S(y,y) + (y,1) \right) \cdot z(i)$.

By assumption, $S(y,z) = \langle y,1 \rangle \cdot \langle z,1 \rangle + \langle y, z \rangle = 0$. Note that this also implies $\langle y, z \rangle \cdot \langle y, z \rangle + \langle y, z \rangle = 0$, implying $\ell(i) = \left( S(y,y) + S(y,z) + S(z,z) + (y,1) \right) \cdot z(i) + \left( S(y,y) + S(y,z) + S(z,z) + (z,1) \right) \cdot y(i)$.

Consequently, the linear part of $Q$ may be written as $\sum_i \ell(i)x(i) = \left( S(y,y) + S(y,z) + S(z,z) + (y,1) \right) \cdot \langle x, z \rangle + \left( S(y,y) + S(y,z) + S(z,z) + (z,1) \right) \cdot \langle x, 1 \rangle + \left(S(y,y) + S(y,z) + S(z,z) + (y,1) \right) \cdot \langle x, 1 \rangle$.

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

4.2 2nd derivatives of a degree 3 polynomial

Let $g(x) = \sum_{i<k} a_{i,j,k} x(i)x(j)x(k)$ be a polynomial of degree 3. For directions $y, z \in \mathbb{F}_N$, consider the second derivative $g_{yz} = \sum z_{i,j} y_{j,k} \cdot x(i)x(j)x(k)$. We need to show that the probability of the vector $g_{yz}$ falling in the affine space $AF_{yz} = yz + Span(y, z, 1)$ is exponentially small.

First, some notation. For $1 \leq i \leq N$, let $G_i$ be a symmetric $N \times N$ matrix over $\mathbb{F}$ with $(G_{j})_{j,k} = (G_{i})_{k,j} = a_{i,j,k}$ 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 $(G_{i})_{j,k} = (G_{i})_{k,j} = (G_{k})_{j,i}$. These matrices are relevant because they describe the vector $g_{yz}$.

**Lemma 18.** $\langle v_{g_{yz}}(i) \rangle = \text{coeff}_{G_i}(g_{yz}, x(i)) = \langle y, G_i \rangle$ (1) An alternative representation of $v_{g_{yz}}$ will be more convenient for us. For $z \in \mathbb{F}_N$, let $G(z) = \sum_{i=1}^N z(i)G_i$. Then $v_{g_{yz}} = G(z) \cdot y$.

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

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

We conclude the proof by a technical claim.

**Proposition 19.** Let $\{A_i\}, i = 1 \ldots N$ be a family of symmetric $N \times N$ matrices over $\mathbb{F}$ with $A_i(k,k) = 1$. Then, for $y, z$ uniformly at random and independently from $\mathbb{F}_N$,

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

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

**Lemma 20.** Let matrices $\{A_i\}$ be as in the proposition. Let $C$ be any fixed symmetric $N \times N$ matrix. Then

$$Pr_{\{z\}}\left\{ \text{rank}(A(z) + C) \leq k - 1 \right\} \leq \left( \frac{3}{4} \right) \cdot \frac{1}{2^N} \cdot \sum_{i=0}^{k-1} \left( \frac{N}{i} \right)$$

**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 $\Pi_{I \subseteq J} z_i$ in $f_I(z)$ is 1. If this is true, deg($f_I - \prod_{i \in J} z_i$) $\leq k - 1$, and the claim of the lemma will follow from Lemma 20.

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

$$\text{det}(B(z)) = \sum_{\sigma \in S_N, \sigma \neq \text{id}} \prod_{i=1}^{N} B_{\sigma(i)}(z) = \sum_{\sigma \in S_N, \sigma \neq \text{id}} \prod_{i=1}^{N} \prod_{i \in \sigma(i)} (z_i + C_{i,i}) \cdot \prod_{\{i < \sigma(i)\}} B_{\sigma(i)}(z) = \prod_{i=1}^{N} (z_i + \text{lower order terms})$$

In the second equality we use the identity $B_{\sigma(i)}(z) = B_{\sigma(i)}(z)$ in $\mathbb{F}$. □

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

Let $p(z) = \text{Pr}_{y,z}\left\{ \langle A(z) \cdot y = y + z + 1 \rangle \right\}$. Clearly $p(z) \leq 2^{-\text{rank}(A(z) + 1)}$.

By Lemma 20, $Pr_{y,z}\left\{ \langle A(z) \cdot y = y + z + 1 \rangle \right\} = E_{p}e_{p} \leq E_{p}2^{-\text{rank}(A(z) + 1)} \leq \frac{1}{N} \cdot \sum_{k=1}^{N} \binom{N}{k} 2^{-k} = \left( \frac{3}{4} \right)^N$.
This concludes the proof of the proposition, and of Theorem 2.

5. PROOF OF LEMMA 3

We need to prove that $S_{a,b}$ over $\mathbb{F}_p$ cannot be approximated by lower degree polynomials. We follow similar approach to the one taken by Green and Tao, with some modifications.

Green and Tao prove ([8]) that $S_a$ over $\mathbb{F}_2$ cannot be approximated by cubic polynomials by using an argument due to Alon and Beigel, which states that any multilinear approximating polynomial can be converted into a symmetric multilinear polynomial (over fewer variables). Since over $\mathbb{F}_2$ all polynomials are multilinear, this leaves to prove that $S_1$ cannot be approximated by a symmetric polynomial, i.e. a polynomial of the form $a_3S_3(x) + a_2S_2(x) + a_1S_1(x) + a_0$. This can be verified by hand.

The first obstacle to using the Alon-Beigel argument in our case is that the approximating polynomial doesn’t have to be multilinear. We overcome this by proving that if $S_{a,b,N}$ has an approximating polynomial of lower degree, then $S_{a,b,N'}$ has a multilinear approximating polynomial over the set $\{0,1\}^{N'}$, for $N' = \Omega(N)$. Now the Alon-Beigel argument can be applied, as in the proof of Green and Tao, to show that there is a symmetric multilinear polynomial which approximates $S_{a,b,N''}$, where $N''$ is in the order of the $p^2$-iterated log of $N'$. We then prove that $S_{a,b,N''}$ cannot be approximated by a linear combination of $\{S_n, N\}_{n < p^2}$, using Lucas’ theorem.

We start by showing that if a polynomial $g$ has correlation $\delta$ with $S_{a,b,N}$, then there is a multilinear polynomial $h$ having correlation $O(\delta)$ with $S_{a,b,N'} (N' = \Omega(N))$ over the set $\{0,1\}^{N'}$ (a polynomial is multilinear if the individual degrees of all the variables are at most 1)

We define a partition of $\mathbb{F}_p^N$, fixing values which are not in $\{0,1\}$. For $I \subseteq [N]$ and values $a = (a_i)_{i \in I} \in \mathbb{F}_p$ define:

$$X_{I,a} = \{(x_1, \ldots, x_N) : \forall i \in I, x_i \in \{0,1\}, \forall j \not\in I, x_j = a_j\}$$

The collection $\{X_{I,a}\}_{I \subseteq [N]}$ forms a partition of $\mathbb{F}_p^N$. Let $\delta_I$ be the average correlation of $S_{a,b,N}$ and $g$ over $X_{I,a}$, i.e.

$$\delta_I,a = \mathbb{E}_{x \in X_{I,a}} e(S_{a,b,N}(x) - g(x))$$

The weighted average of $\delta_I,a$ is at least $\delta$, where the weight of $\delta_I,a$ is $2^{|I|}p^{-N}$. It is easy to see that here must exist a pair $(I,a)$ with $|I| \leq 2^N \delta$ s.t. $\delta_I,a = \Omega(\delta)$.

The restriction of $S_{a,b,N}$ to $X_{I,a}$ is a symmetric polynomial with the degree-$p^2$ part being $S_{a,b}[[I]]$. The restriction of $g$ to $X_{I,a}$ may be non-multilinear, but since it is evaluated only on points in $\{0,1\}^{|I|}$, it can be replaced by a multilinear polynomial of the same degree which agrees with it on $\{0,1\}^{|I|}$. Our new polynomial $h$ will be this multilinear polynomial, plus the lower degree terms of the restriction of $S_{a,b,N}$.

Now we have a multilinear polynomial $h$ approximating $S_{a,b,N''}$. We can now use the Alon-Beigel argument ([1]), which uses Ramsey-type reasoning to gradually replace $h$ by another multilinear symmetric polynomial $k$, now on $N''$ variables, which approximates $S_{a,b,N''}$ with the same correlation. $N''$ is of the order of $p^2$-iterated log of $N'$.

To finish, we need to prove that $S_{a,b,N''}$ cannot be approximated on $\{0,1\}^{N''}$ by a lower degree multilinear symmetric polynomial, i.e. by a polynomial of the form $\sum_{n=0}^{l} a_n S_n(x)$. In order to do so, we use Lucas’ theorem, which we now state.

**THEOREM 21 (LUCAS’ THEOREM).** Let $p$ be a prime number. Let $a, b$ be two natural numbers, and consider their development in base $p$, i.e. $a = a_0 + a_1p + a_2p^2 + \ldots$ and $b = b_0 + b_1p + b_2p^2 + \ldots$, where $0 \leq a_i, b_i \leq p - 1$. Then:

$$\left(\frac{a}{b}\right) \equiv \left(\frac{a_0}{b_0}\right) \left(\frac{a_1}{b_1}\right) \left(\frac{a_2}{b_2}\right) \ldots \pmod{p}$$

Consider $x \in \{0,1\}^{N''}$, and let $w(x) = x_1 + x_2 + \ldots + x_N''$ be the number of 1’s in $x$. It’s easy to see that $S_n(x) = (w(x))^{n}(\pmod{p})$. Let $w(x) = w_0(x) + w_1(x)p + w_2(X)p^2 + \ldots$ be the development of $w(x)$ in base $p$. According to Lucas’ theorem, $S_{a,b}(x) = w_2(x)$ while $S_n(x)$ for $n < p^2$ depends only on $w_0(x)$ and $w_1(x)$. To conclude the proof, notice that for $N \rightarrow \infty$, the joint distribution of $(w_0(x), w_1(x), w_2(x))$ is uniform over $\mathbb{F}_p^3$.

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

6. ACKNOWLEDGEMENTS

We are grateful to Ben Green and Terence Tao for sending us their paper and by bringing the argument of Alon and Beigel to our attention. The first author thanks his supervisor, Omer Reingold, for useful comments and for his constant support and interest in the work.

7. 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] N. Alon, T. Kaufman, M. Krivelevich, S. Lifschitz, D. Ron, Testing low-degree polynomials over GF(2), RANDOM-APPROX 2003, pp. 188-199.

[3] M. Bellare, D. Coppersmith, J. Hastad, M. Kiwi, M. Sudan, Linearity testing in characteristic 2, IEEE Trans. Inform. Theory, vol. IT-42, 6, 1996, 1782-1795.

[4] M. Blum, M. Luby, R. Rubinfeld, Self-testing/correcting with applications to numerical problems, J. Comp. Sys. Sci., 47, 3, 1993.

[5] O. Goldreich, S. Goldwasser, and D. Ron, Property Testing and its connection to learning and approximation, J. ACM, 45(4), 1998, pp. 653-750.

[6] W. T. Gowers, A new proof of Szemerédi’s theorem, GAFA Vol. 11(2001), pp. 465-588.

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

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

[9] B. Host, B. Kra, Nonconventional ergodic averages and nilmanifolds, Annals of Math. 161, 1 (2005) 397-488.

[10] T. Kaufman, D. Ron, Testing polynomials over general fields, FOCS 2004, pp. 413-422.

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

[12] A. Samorodnitsky, Low degree tests at large distances, STOC ’07.

[13] T. Tao, Structure and randomness in combinatorics, FOCS ’07.

556