THEORY OF COMPUTING, Volume 7 (2011), pp. 131–145
*www.theoryofcomputing.org*

# Inverse Conjecture for the Gowers Norm is False

Shachar Lovett[*] Roy Meshulam[†] Alex Samorodnitsky[‡]

*Received: October 8, 2008; revised: April 29, 2011; published: July 12, 2011.*

**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 $f$ is non-trivially correlated to 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 $d = 4$, by presenting an explicit function whose 4-th Gowers norm is non-negligible, but whose correlation to any polynomial of degree 3 is exponentially small. Essentially the same result (with different correlation bounds) was independently obtained by Green and Tao (2009).

**ACM Classification:** F.2.2

**AMS Classification:** 05E99

**Key words and phrases:** Inverse Gowers conjecture, additive combinatorics, Gowers norm

## 1 Introduction

We consider multivariate functions over finite fields. The main question of interest here is whether these functions can be non-trivially approximated by a low-degree polynomial. Fix a prime $p$. Let $\mathbb{F} = \mathbb{F}_p$ denote the finite field with $p$ elements. Let $\omega = e^{2\pi i/p}$ be the primitive $p$-th root of unity. Denote by $e(x)$ the exponential function taking $x \in \mathbb{F}$ to $\omega^x \in \mathbb{C}$. For two functions $f, g : \mathbb{F}^N \to \mathbb{F}$, let

$$\langle f, g \rangle := \mathbb{E}_x \left[ e(f(x) - g(x)) \right] .$$

[*] Supported by DMS grant 0835373.
[†] Supported by ISF and BSF grants
[‡] Supported by ISF grant 039-7682 and by BSF grant 2006377.

DOI: 10.4086/toc.2010.v007a009

A function $f$ is non-trivially approximable by a degree-$d$ polynomial if

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

for some polynomial $g(x)$ of degree at most $d$ over $\mathbb{F}$. More precisely, in this paper 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, where the remaining parameters, that is the field size $p$, the degree $d$ and the offset $\varepsilon$ are fixed, independent of $N$.

**Definition 1.1.** Fix a finite field $\mathbb{F} = \mathbb{F}_p$. A sequence of functions $\{f_N : \mathbb{F}^N \to \mathbb{F}\}$ is non-trivially approximable by degree-$d$ polynomials if there exists a sequence of degree-$d$ polynomials $\{g_N : \mathbb{F}^N \to \mathbb{F}\}$ and an offset $\varepsilon > 0$ such that for all $N$,

$$|\langle f_N, g_N\rangle| > \varepsilon\,.$$

A counting argument shows that a generic function cannot be approximated by a polynomial of low degree. The problem of showing a *specific* given function to have no non-trivial approximation by low-degree polynomials has been extensively investigated, since solutions to this problem have many applications in complexity (cf. discussion and references in [1, 3, 16], as well as an excellent survey by Viola on correlation bounds [15]).

This paper studies a technical tool that measures distance from low-degree polynomials. This is the Gowers norm, introduced in [4]. For a function $f : \mathbb{F}^N \to \mathbb{F}$ 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 $y_1, \ldots, y_k \in \mathbb{F}^n$ we take the iterated derivative in these directions to be

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

It is easy to see that this definition does not depend on the ordering of $y_1, \ldots, y_k$. The "$k$-th Gowers norm" $\|f\|_{U^k}$ of $f$ is

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

More accurately, as shown in [4], 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 that $|\langle f, g\rangle| > \varepsilon$ for $g$ of degree at most $d$, implies $\|f\|_{U^{d+1}} > \varepsilon$ [5]. That is to say, if $f$ is non-trivially close to a degree-$d$ polynomial, this can be detected via an appropriate Gowers norm.

This discussion naturally leads to the inverse conjecture [5, 9, 11], that is, if the $(d+1)$-th Gowers norm of $f$ is non-trivial, then $f$ is non-trivially approximable by a degree-$d$ polynomial. This conjecture is referred to as the *Inverse Conjecture for the Gowers Norm (ICGN)*. This conjecture is easily seen to hold for $d = 1$ and has been proved also for $d = 2$ [5, 9]. It is of interest to prove this conjecture for higher values of $d$. As the conjecture remained open, special cases of the inverse conjecture were also studied (i. e., the $d$-vs-$(d-1)$ conjecture [3], and the inverse conjecture for low-degree polynomials [6]).

Our main result is that ICGN is false. This note is aimed at providing a self-contained proof for the case of $p = 2$. The result for general $p$ can be found in the conference version [7], where the bounds obtained are much weaker than the ones presented in this note.

From now on we consider functions $f : \mathbb{F}_2^N \to \mathbb{F}_2$. For $x \in \mathbb{F}_2^N$ let $x(i)$ denote the $i$-th coordinate of $x$. Let $S_4$ denote the symmetric polynomial of degree 4 in $N$ variables,

$$S_4(x) = \sum_{1 \le i < j < k < \ell \le N} x(i)x(j)x(k)x(\ell)\,.$$

We prove two results, whose combination shows that ICGN is false over $\mathbb{F}_2$.

**Theorem 1.2.** *There exists an absolute constant $c > 0$ such that*

$$\|S_4\|_{U^4} \ge c\,.$$

**Theorem 1.3.** *There exists an absolute constant $\alpha < 1$ such that for any cubic polynomial $g(x)$ over $\mathbb{F}_2$*

$$\langle S_4, g\rangle \le \alpha^N\,.$$

## 1.1 Related work

Our results have a large overlap with a recent work of Green and Tao [6].

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 $p$. 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 $d = 4$.

The calculation of the 4-th Gowers norm of $S_4$ in this paper as well as in [6] follows by a direct calculation of the bias of the 4-th derivative polynomial of $S_4$, where the analysis in [6] is somewhat simpler and cleaner. In any case, this is the simpler and more direct part of the proof in both papers.

The proof of non-approximability of $S_4$ by lower-degree polynomials in [6] 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 analysis gives 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, as it can be readily extended to the case of a general prime $p$.

## 1.2 The case of a general prime field

Let $\mathbb{F} = \mathbb{F}_p$. Let $S_d$ denote the symmetric polynomial of degree $d$ in $N$ variables,

$$S_d(x) = \sum_{S \subset [N], |S| = d} \prod_{i \in S} x(i)\,.$$

The polynomial $S_{p^2}$ over $\mathbb{F}_p$ provides a counterexample for ICGN over $\mathbb{F}_p$. One can prove the following:

1. For any $d \geq 2p$, there exists an absolute constant $\varepsilon_d > 0$, such that for any $N$
$$\|S_d\|_{U^d} \geq \varepsilon_d\,.$$
2. For any polynomial $g$ of degree at most $p^2-1$,
$$\langle S_{p^2}, g\rangle \leq \left(\log^{(p^2)} N\right)^{-1} = o_N(1)$$
where $\log^{(t)}$ denotes the $t$-fold iteration of the logarithm function.

The proof of the first claim entails calculating and working with the iterated derivatives of $S_d$. It is similar in the outline to the argument for $S_4$ over $\mathbb{F}_2$, but requires some additional technicalities. The proof of the second claim follows by an appropriate adaptation of the argument of Alon and Beigel.

### 1.3 Subsequent work

Subsequent to this work, Bergelson, Tao, and Ziegler [2, 13, 14] proved a refined version of ICGN. Let $F : \mathbb{F}_p^N \to \mathbb{C}$ be a function. The derivative of $F$ in direction $y \in \mathbb{F}_p^N$ is defined as $F_y(x) = F(x+y)\overline{F(x)}$, and iterated derivatives are defined analogously. Note that when $F$ takes values which are $p$-roots of unity, i. e., when $F(x) = e^{f(x)2\pi i/p}$ for some $f : \mathbb{F}_p^N \to \mathbb{F}_p$, then this definition of derivatives coincides with our previous definition, i. e., $F_y(x) = e^{f_y(x)2\pi i/p}$.

A function $F : \mathbb{F}_p^N \to \mathbb{C}$ is said to be a *non-classical polynomial* of degree $d$ if for any $y_1, \dots, y_{d+1} \in \mathbb{F}_p^N$ we have $F_{y_1,\dots,y_{d+1}} \equiv 1$. Clearly, if $f : \mathbb{F}_p^N \to \mathbb{F}_p$ is a degree $d$ polynomial, then $F = e^{f(x)2\pi i/p}$ is a non-classical polynomial of degree $d$. However, there exist other examples of non-classical polynomials. Let $f(x)$ be a degree $d-(p-1)(\ell-1)$ polynomial over $\mathbb{Z}_{p^\ell}$. Then $F(x) = e^{f(x)2\pi i/p^\ell}$ is also a non-classical polynomial of degree $d$ (note that $F$ is still evaluated over $\mathbb{F}_p^N$). In fact, this is a complete classification of all non-classical polynomials [12].

**Theorem 1.4** ([2, 13, 14])**.** *Fix prime $p$, $d \geq 1$ and $\varepsilon > 0$. Let $F : \mathbb{F}_p^N \to \mathbb{C}$ be a function such that $\|F\|_\infty \leq 1$ and $\|F\|_{U^{d+1}} \geq \varepsilon$. Then there exists a non-classical polynomial $G : \mathbb{F}_p^N \to \mathbb{C}$ of degree $d$ such that*
$$|\langle F, G\rangle| = \left|\mathbb{E}_{x\in\mathbb{F}_p^N}\left[F(x)\overline{G(x)}\right]\right| \geq \delta$$
*where $\delta = \delta(\mathbb{F}_p, d, \varepsilon) > 0$, i. e., $\delta$ does not depend on $N$.*

Consider, in this framework, the counterexample of $S_4$ over $\mathbb{F}_2$ discussed in this paper. For $x \in \mathbb{F}_2^n$ and $f(x) = x(1) + \cdots + x(n) \pmod 8$, i. e., $f$ is a linear polynomial over $\mathbb{Z}_8$, observe that $S_4(x)$ is the most significant bit of $f(x)$. Moreover, it can be easily checked that $(-1)^{S_4}$ is correlated to $F(x) = e^{\frac{2\pi i}{8}f(x)}$, a non-classical polynomial of degree 3 (to verify, note that both functions depend only on the hamming weight of $x$ modulo 8). Thus, $S_4$ has a noticeable 4-th Gowers norm, and is indeed correlated to a non-classical cubic polynomial.

## 2 $S_4$ has high 4-Gowers norm

This section contains the proof of Theorem 1.2. We show that there exists a positive constant $c > 0$ such that $\|S_4\|_{U^4} \geq c$, independent of the number of variables $N$.

We start by explicitly describing the 4-th iterative derivative of $S_4$.

**Claim 2.1.** *Let $S(y_1, y_2, y_3, y_4) = \sum_{i_1 \neq i_2 \neq i_3 \neq i_4} y_1(i_1) y_2(i_2) y_3(i_3) y_4(i_4)$, where the sum is over distinct elements $i_1, i_2, i_3, i_4 \in [N]$. Then*

$$(S_4)_{y_1,y_2,y_3,y_4}(x) = S(y_1, y_2, y_3, y_4) .$$

*In particular,*

$$\|S_4\|_{U^4}^{16} = \mathbb{E}_{y_1,y_2,y_3,y_4}[(-1)^{S(y_1,y_2,y_3,y_4)}] .$$

*Proof.* Let $i_1, i_2, i_3, i_4$ be distinct elements of $[N]$. Consider a monomial $m(x) = x(i_1)x(i_2)x(i_3)x(i_4)$. Its 4-th iterated derivative in directions $y_1, \ldots, y_4$ is, by definition,

$$m_{y_1,y_2,y_3,y_4}(x) = \sum_{I \subseteq [4]} \prod_{j=1}^{4} \left( x(i_j) + \sum_{k \in I} y_k(i_j) \right) .$$

Expanding the right hand side as a sum of monimials, we observe that the only monomials appearing an odd number of times are of the form $\prod_{j=1}^{4} y_j(i_{\pi(j)})$, where $\pi$ is a permutation on 4 elements. Since we are working in $\mathbb{F}_2$, we conclude

$$m_{y_1,y_2,y_3,y_4}(x) = \sum_{\pi \in \mathrm{Sym}_4} \prod_{j=1}^{4} y_j(i_{\pi(j)}) .$$

The claim now follows by summing over all monomials in $S_4$. □

Let $M = M(y_1, y_2, y_3, y_4)$ denote the $4 \times N$ matrix over $\mathbb{F}_2$ whose rows are given by $y_1, y_2, y_3, y_4$. Let $M_{i_1,i_2,i_3,i_4}$ denote the $4 \times 4$ minor of $M$ restricted to columns $i_1, i_2, i_3, i_4$. Observe that $S(y_1, y_2, y_3, y_4)$ can be expressed as the sum over all permanents of $4 \times 4$ minors of $M$.

**Claim 2.2.** $S(y_1, y_2, y_3, y_4) = \sum_{i_1 < i_2 < i_3 < i_4} \mathrm{Per}(M_{i_1,i_2,i_3,i_4})$.

We recall the Binet-Cauchy formula.

**Lemma 2.3.** *Let $A$ be an $m \times n$ matrix with $m \leq n$ over a field. For $I \subset [n]$ such that $|I| = m$, let $A_I$ be the $m \times m$ minor of $A$ restricted to columns of $I$. Then*

$$\sum_{I \subset [n], |I| = m} \mathrm{Det}(A_I)^2 = \mathrm{Det}(AA^t) .$$

Recall that $x^2 = x$ in $\mathbb{F}_2$, and also that determinants and permanents are identical in $\mathbb{F}_2$. Therefore we may apply the Binet-Cauchy formula for $M$, using Claim 2.2, and conclude that

**Corollary 2.4.** $S(y_1, y_2, y_3, y_4) = \mathrm{Det}(MM^t)$.

*Proof.*

$$S(y_1,y_2,y_3,y_4) = \sum_{I\subset[n],|I|=4} \mathrm{Per}(M_I) = \sum_{I\subset[n],|I|=4} \mathrm{Det}(M_I) = \sum_{I\subset[n],|I|=4} \mathrm{Det}(M_I)^2 = \mathrm{Det}(MM^t)\,.$$

□

The matrix $MM^t$ is a $4\times 4$ symmetric matrix, whose $(i,j)$ entry is given by $\langle y_i,y_j\rangle$. We next show that the distribution of $MM^t$ where $y_1,y_2,y_3,y_4$ are chosen uniformly from $\mathbb{F}_2^N$ is very close to the distribution of a uniformly chosen $4\times 4$ symmetric matrix. Recall that the statistical distance between two random variables $X,Y$ is given by

$$\mathrm{sd}(X,Y) = \frac{1}{2}\sum_a |\Pr[X=a]-\Pr[Y=a]|\,.$$

**Claim 2.5.** *Let $X$ denote a uniform $4\times 4$ symmetric matrix over $\mathbb{F}_2$. Then the distribution of $MM^t$ for uniformly chosen $y_1,y_2,y_3,y_4\in\mathbb{F}_2^N$ is $O(2^{-N})$ close to the distribution of $X$ (in statistical distance). In particular*

$$\Pr[\mathrm{Det}(MM^t)=0] \ge \Pr[\mathrm{Det}(X)=0] - O(2^{-N})\,.$$

The proof of the claim uses the following fact, which is easily verified using standard Fourier analysis in $\mathbb{F}_2^N$ (see, e. g., [3]).

**Claim 2.6.** *Let $Y\in\mathbb{F}_2^k$ be a random variable, such that for all non-zero $c\in\mathbb{F}_2^k$*

$$\left|\Pr[\langle Y,c\rangle = 0] - \frac{1}{2}\right| \le \alpha\,.$$

*Then the distribution of $Y$ is $2^k\cdot\alpha$ close (in statistical distance) to the uniform distribution over $\mathbb{F}_2^k$.*

*Proof of Claim 2.5.* We can consider the symmetric matrix $MM^t$ as a vector in $\mathbb{F}_2^{10}$ indexed by $\{(i,j): 1\le i\le j\le 4\}$. Let $c\in\mathbb{F}_2^{10}$ be a non-zero vector. We will prove that

$$\left|\Pr[\langle MM^t,c\rangle = 0] - \frac{1}{2}\right| \le O\left(2^{-N}\right)\,.$$

This will imply that the distribution of $MM^t$ is $O(2^{-N})$ close to the uniform distribution over symmetric $4\times 4$ matrices. To prove this, for $a_1,a_2,a_3,a_4\in\mathbb{F}_2$ define the $4\times 4$ symmetric matrix $A(a_1,a_2,a_3,a_4)$ whose $(i,j)$ entry is $a_ia_j$. Note that $MM^t = \sum_{i=1}^N A(y_1(i),y_2(i),y_3(i),y_4(i))$. Therefore,

$$2\Pr[\langle MM^t,c\rangle = 0]-1 = \mathbb{E}\left[(-1)^{\langle MM^t,c\rangle}\right] = \prod_{i=1}^N \mathbb{E}\left[(-1)^{A(y_1(i),y_2(i),y_3(i),y_4(i))}\right]\,.$$

Thus we have

$$2\Pr[\langle MM^t,c\rangle = 0]-1 = \left(\mathbb{E}_{a_1,a_2,a_3,a_4\in\mathbb{F}_2}\left[(-1)^{\langle A(a_1,a_2,a_3,a_4),c\rangle}\right]\right)^N\,.$$

To conclude, note that $\langle A(a_1,a_2,a_3,a_4),c\rangle$ is a nonzero quadratic polynomial over $\mathbb{F}_2$ in the variables $a_1,a_2,a_3,a_4$. Standard bounds on the number of roots of polynomials over a field [10] imply

$$\frac{1}{4} \le \Pr[\langle A(a_1,a_2,a_3,a_4),c\rangle = 0] \le \frac{3}{4}\,.$$

Therefore

$$\left|2\Pr\left[\langle MM^t,c\rangle = 0\right] - 1\right| \le 2^{-N}$$

which, by Claim 2.6, gives

$$\mathrm{sd}(MM^t,X) \le 1024\cdot 2^{-N}\,.$$

□

To conclude the proof of Theorem 1.2, we use the following fact, which can be easily verified.

**Claim 2.7.** *Let $X$ be a uniformly chosen $4\times 4$ symmetric matrix over $\mathbb{F}_2$. Then* $\Pr[\mathrm{Det}(X)=0] = \frac{9}{16}$.

We now conclude the proof of Theorem 1.2. Combining Claims 2.1, 2.4, 2.5 and 2.7,

$$\|S_4\|_{U^4}^{16} = \mathbb{E}\left[(-1)^{S(y_1,y_2,y_3,y_4)}\right] = \mathbb{E}\left[(-1)^{\mathrm{Det}(MM^t)}\right] \ge \mathbb{E}\left[(-1)^{\mathrm{Det}(X)}\right] - O\left(2^{-N}\right) \ge \frac{1}{8} - O\left(2^{-N}\right)\,.$$

# 3 $S_4$ has no correlation with cubics

This section contains the proof of Theorem 1.3. We show there is an absolute constant $\alpha > 0$ such, that for any cubic polynomial $g$ 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 3.1.** *For any two functions $f$ and $g$ holds*

$$\langle f,g\rangle^4 \le \mathbb{E}_y[\langle f_y,g_y\rangle^2]\,.$$

*Proof.* This is an immediate corollary of a lemma in [9], but we give the elementary proof for completeness. By the Cauchy-Schwarz inequality,

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

□

**Corollary 3.2.**

$$\langle f,g\rangle^8 \le \mathbb{E}_{y,z}[\langle f_{y,z},g_{y,z}\rangle^2]\,.$$

We will show that for any polynomial $g$ of degree at most 3 it holds that

$$\mathbb{E}_{y,z}\left[\left\langle (S_4)_{y,z}, g_{y,z}\right\rangle^2\right] \le \exp(-\alpha N)\,.$$

First, here is a brief overview of the argument.

The advantage in taking second derivatives is that a second derivative of $g$ is a linear function, and a second derivative of $S_4$ is a quadratic. Fix $y,z$ and let $L(x) = g_{y,z}(x)$ and $Q(x) = (S_4)_{y,z}(x)$ be the corresponding linear and quadratic function. The correlation between $L$ and $Q$ is the absolute value of the Fourier coefficient of $Q$ which corresponds to the character given by $L$. The Fourier spectrum of quadratic functions is well understood using a theorem of Dixon. It turns out that given the choice of directions $y,z$, the quadratic $Q$ falls into one of two classes. For half of the choices for $y,z$, all Fourier coefficients of $Q$ are exponentially small, and hence in particular the correlation between $Q$ and $L$ is exponentially small. For the remaining choices of $y,z$ however, the quadratic function $Q$ has only a constant number of Fourier coefficients. However, they all lie in an explicitly given 3-dimensional affine subspace depending on $y,z$. We then argue that for any fixed cubic polynomial $g$, 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 derivatives of $S_4$.

## 3.1 Second derivatives of $S_4$

Fix directions $y,z \in \mathbb{F}_2^n$, and let $Q(x) = (S_4)_{y,z}(x)$. Write

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

We start by computing the coefficients $q_{i,j}$. Let $\mathcal{S}(y,z) = \sum_{k\neq\ell} y(k)z(\ell) = \langle y,z\rangle + \langle y,\mathbf{1}\rangle \cdot \langle z,\mathbf{1}\rangle$, where $\mathbf{1} \in \mathbb{F}_2^N$ denotes the all-1 vector.

**Claim 3.3.**

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

*Proof.* Direct computation gives

$$q_{i,j} = \sum_{\substack{k\neq\ell \\ k,\ell\notin\{i,j\}}} y(k)z(\ell)\,.$$

Using the inclusion-exclusion formula yields

$$\begin{aligned} q_{i,j} &= \sum_{k\neq\ell} y(k)z(\ell) - \sum_{\substack{k\in\{i,j\} \\ \ell\notin\{i,j\}}} y(k)z(\ell) - \sum_{\substack{\ell\in\{i,j\} \\ k\notin\{i,j\}}} y(k)z(\ell) + \sum_{\{k,\ell\}=\{i,j\}} y(k)z(\ell) \\ &= \mathcal{S}(y,z) - (y(i)+y(j))\big(\langle z,\mathbf{1}\rangle - z(i) - z(j)\big) \\ &\qquad - (z(i)+z(j))\big(\langle y,\mathbf{1}\rangle - y(i) - y(j)\big) + \big(y(i)z(j)+y(j)z(i)\big). \end{aligned}$$

Rearranging, and using the fact that the computation is over $\mathbb{F}_2$, we obtain the claim. □

At this point we invoke (a corollary of) a theorem of Dixon (see for example [8], Section 15, Theorem 5):

**Theorem 3.4.** *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 $B$ with zeros on the diagonal and off-diagonal entries given by $B_{i,j} = B_{j,i} = q_{i,j}$. Let the rank of $B = 2h$ (it is always even). Then the function $(-1)^Q$ has exactly $2^{2h}$ non-zero Fourier coefficients all 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 each 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 3.4, the Fourier coefficients of $(-1)^Q$ are exponentially small. In fact,

**Corollary 3.5.** *If $\mathcal{S}(y,z) = 1$ then, for any cubic polynomial $g(x)$,*

$$\big\langle (S_4)_{y,z}, g_{y,z} \big\rangle \le 128 \cdot 2^{-N}.$$

*Proof.* The matrix $B$ has rank at least $N-7$, since $J$ has rank at least $N-1$ and each matrix of the form $u \otimes v$ is a rank one matrix. The claim now follows from Theorem 3.4. □

Recall that we wish to show that for any polynomial $g$ of degree at most 3, the average value of $\langle (S_4)_{y,z}, g_{y,z}\rangle$ for uniformly chosen $y,z \in \mathbb{F}_2^n$ is exponentially small in $n$. Corollary 3.5 shows that this holds for any $y,z$ whenever $\mathcal{S}(y,z) = 1$. We, therefore, may assume that $\mathcal{S}(y,z) = 0$ from now on. In this case the quadratic part of $Q$ may be simplified as

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

Here $yz$ denotes the *pointwise product* of the vectors $y$ and $z$, that is $(yz)(i) = y(i)z(i)$.

Observe, that the above computation implies the non-zero Fourier coefficients of $\sum_{i<j} q_{i,j} x(i)x(j)$ lie in an affine subspace of $\mathbb{F}_2^n$ of dimension at most 3, given by $yz + \mathrm{Span}(\mathbf{1}, y, z)$.

Next, consider the linear part $\langle \ell, x\rangle = \sum_i \ell(i)x(i)$ of $Q$.

**Claim 3.6.** *If $\mathcal{S}(y,z) = 0$ then $\ell \in \mathrm{Span}(y,z,\mathbf{1})$.*

*Proof.* Straightforward calculation gives that

$$\ell(i) = \sum_{j<k<l\neq i} \big(y(k)y(l)z(j) + y(j)y(l)z(k) + y(j)y(k)z(l)\big) + \big(y(j)z(k)z(l) + y(k)z(j)z(l) + y(l)z(j)z(k)\big).$$

This can be directly verified to be equal to

$$\begin{aligned}\big(\mathcal{S}(y,z) + \mathcal{S}(z,z) + \langle z, \mathbf{1}\rangle\big) \cdot y(i) + \big(\mathcal{S}(y,z) + \mathcal{S}(y,y) + \langle y, \mathbf{1}\rangle\big) \cdot z(i) \\ + \big(\mathcal{S}(y,y)\cdot\langle z, \mathbf{1}\rangle + \mathcal{S}(z,z)\cdot\langle y, \mathbf{1}\rangle + \langle y, z\rangle \cdot \langle y+z, \mathbf{1}\rangle\big).\end{aligned}$$

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

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

□

Consequently, we deduce

**Corollary 3.7.** *If $\mathcal{S}(y,z) = 0$ then 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* $\mathrm{AF}_{y,z} = yz + \mathrm{Span}\,(y,z,\mathbf{1})$.

### 3.2 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 $y,z \in \mathbb{F}_2^N$, 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 $\mathrm{AF}_{y,z} = yz + \mathrm{Span}\,(y,z,\mathbf{1})$ is exponentially small.

First, we fix some notation. For $1 \le i \le N$, let $G_i$ be a symmetric $N \times N$ matrix over $\mathbb{F}_2$ 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 3.8.**

- $v_{y,z}(i) = \mathrm{coef}_{x(i)}\,(g_{y,z}(x)) = \langle y, G_i z\rangle$.
- *An alternative representation of $v_{y,z}$ will be more convenient for us. For $z \in \mathbb{F}_2^N$, 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 $\{v_{y,z} \in \mathrm{AF}_{y,z}\}$. 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 $A_i = G_i + e_i \otimes e_i$, $i = 1,\dots,N$. 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 3.9.** *Let $\{A_i\}$, $i = 1,\dots,N$ be a family of symmetric $N\times N$ matrices over $\mathbb{F}_2$ with $A_i(k,k) = \delta_{ik}$. Then, for $y,z$ uniformly at random and independently from $\mathbb{F}_2^N$,*

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

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

**Lemma 3.10.** *Let matrices $\{A_i\}$ be as in the proposition. Let $C$ be any fixed symmetric $N\times N$ matrix. Then*

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

The proof of Lemma 3.10 uses the following estimate on the number of common zeros of a set of polynomials.

**Lemma 3.11.** *Let $\{f_I\}$ be a set of $K = \binom{N}{k}$ polynomials over $\mathbb{F}_2$, indexed by k-subsets $I$ of $[N]$. Assume that for any such subset $I$ holds*

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

*Then,*

$$\Pr_{x\in\mathbb{F}_2^n}[f_1(x) = \cdots = f_K(x) = 0] \le \frac{1}{2^N}\sum_{j=0}^{k-1}\binom{N}{j}.$$

We defer the proof of Lemma 3.11 to Subsection 3.3.

*Proof of Lemma 3.10.* Consider a family of $\binom{N}{k}$ polynomials $f_I$ on $\mathbb{F}_2^N$. 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 3.11.

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

$$\begin{aligned}\mathrm{Det}(B(z)) &= \sum_{\substack{\sigma\in S_N\\ \sigma=\sigma^{-1}}}\prod_{i=1}^N B_{i\sigma(i)}(z) = \sum_{\substack{\sigma\in S_N\\ \sigma=\sigma^{-1}}}\prod_{\{i:\sigma(i)=i\}}(z_i + C_{i,i})\cdot\prod_{\{i:i<\sigma(i)\}} B_{i\sigma(i)}(z)\\ &= \prod_{i\in I}^{n} z_i + \text{lower order terms}\end{aligned}$$

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

We now prove Propostion 3.9.

*Proof of Proposition 3.9.* Let $I$ denote the identity $N \times N$ matrix.

Let $p(z) = \Pr_y[A(z)\cdot y = y+z+\mathbf{1}]$. Clearly $p(z) \le 2^{-\operatorname{rank}(A(z)+I)}$. By Lemma 3.10,

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

□

Summing up, for any cubic polynomial $g(x)$,

- For any directions $y,z \in \mathbb{F}_2^n$ such that $\mathcal{S}(y,z) = 1$ we have $\langle (S_4)_{y,z}, g_{y,z}\rangle \le O\left(2^{-N}\right)$.
- For all but exponentially few directions $y,z \in \mathbb{F}_2^n$ such that $\mathcal{S}(y,z) = 0$ we have $\langle (S_4)_{y,z}, g_{y,z}\rangle = 0$.

Hence $\mathbb{E}_{y,z}[\langle (S_4)_{y,z}, g_{y,z}\rangle] = 2^{-\Omega(n)}$, and by Lemma 3.1 we have $\langle S_4, g\rangle = 2^{-\Omega(n)}$. This concludes the proof of Theorem 1.3.

## 3.3 Estimates on the number of common zeroes of some families of polynomials

The main claim of this subsection is the following proposition.

**Proposition 3.12.** *Fix a prime $p$ and let $\mathbb{F} = \mathbb{F}_p$. Let $M$ be the ring of $\mathbb{F}$-valued functions on $\mathbb{F}^N$, that is $M = \mathbb{F}[x_1,\dots,x_N]/I$, where $I$ is the ideal $\left(x_1^p - x_1,\dots,x_N^p - x_N\right)$. Let $f_1,\dots,f_K$ be polynomials in $M$. Let $S$ be the set of common zeroes of $f_1,\dots,f_K$, that is*

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

*Then*

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

*where $J$ is the ideal generated by $\{f_i\}$, and $dim\,(M/J)$ denotes the dimension of $M/J$, 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. □

We now prove Lemma 3.11.

*Proof of Lemma 3.11.* 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 $k-1$. □

## 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. We thank the anonymous reviewers for their careful reading and helpful comments.

## References

[1] NOGA ALON AND RICHARD BEIGEL: Lower bounds for approximations by low degree polynomials over $\mathbb{Z}_m$. In *Proc. 16th Ann. IEEE Conf. Comput. Complexity (CCC)*, pp. 184–187, Washington, DC, USA, 2001. IEEE Comp. Soc. Press. [doi:10.1109/CCC.2001.933885] 132, 133

[2] VITALY BERGELSON, TERENCE TAO, AND TAMAR ZIEGLER: An inverse theorem for the uniformity seminorms associated with the action of $\mathbb{F}_p^\infty$. *Geom. Funct. Anal.*, 19:1539–1596, 2010. [doi:10.1007/s00039-010-0051-1] 134

[3] ANDREJ BOGDANOV AND EMANUELE VIOLA: Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39:2464–2486, April 2010. [doi:10.1137/070712109] 132, 136

[4] TIMOTHY GOWERS: A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001. [doi:10.1007/s00039-001-0332-9] 132

[5] BEN GREEN AND TERENCE TAO: An inverse theorem for the Gowers $U^3(G)$ norm. *Proc. Edinb. Math. Soc. (2)*, 51:73–153, 2008. [doi:10.1017/S0013091505000325] 132

[6] BEN GREEN AND TERENCE TAO: The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009. http://cdm.ucalgary.ca/index.php/cdm/article/view/133/98. 132, 133

[7] SHACHAR LOVETT, ROY MESHULAM, AND ALEX SAMORODNITSKY: Inverse conjecture for the Gowers norm is false. In *Proc. 40th STOC*, pp. 547–556, New York, NY, USA, 2008. ACM Press. [doi:10.1145/1374376.1374454] 133

[8] F. J. MACWILLIAMS AND N. J. A. SLOANE: *The Theory of Error Correcting Codes*. Amsterdam, North-Holland, 1977. 139

[9] ALEX SAMORODNITSKY: Low-degree tests at large distances. In *Proc. 39th STOC*, pp. 506–515, New York, NY, USA, 2007. ACM Press. [doi:10.1145/1250790.1250864] 132, 137

[10] JACOB T. SCHWARTZ: Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4):701–717, 1980. [doi:10.1145/322217.322225] 137

[11] TERENCE TAO: Structure and randomness in combinatorics. In *Proc. 48th FOCS*, pp. 3–15. IEEE Comp. Soc. Press, October 2007. [doi:10.1109/FOCS.2007.17] 132

[12] TERENCE TAO: Some notes on "non-classical" polynomials in finite characteristic, 2008. Online blog: http://terrytao.wordpress.com/2008/11/13/some-notes-on-non-classical-polynomials-in-finite-characteristic. 134

[13] TERENCE TAO AND TAMAR ZIEGLER: The inverse conjecture for the Gowers norm over finite fields via the correspondence principle. *Analysis & PDE*, 3(1):1–20, 2010. [doi:10.2140/apde.2010.3.1] 134

[14] TERENCE TAO AND TAMAR ZIEGLER: The inverse conjecture for the Gowers norm over finite fields in low characteristic. Submitted, 2011. 134

[15] EMANUELE VIOLA: Guest column: Correlation bounds for polynomials over $\{0,1\}$. *SIGACT News*, 40:27–44, February 2009. [doi:10.1145/1515698.1515709] 132

[16] EMANUELE VIOLA AND AVI WIGDERSON: Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4:137–168, 2008. [doi:10.4086/toc.2008v004a007] 132

## AUTHORS

Shachar Lovett
member, School of Mathematics
Institute for Advanced Study, Princeton, NJ
slovett@math.ias.edu
http://www.math.ias.edu/~slovett

Roy Meshulam
professor, Department of Mathematics
The Technion, Haifa, Israel
meshulam@math.technion.ac.il
http://www.math.technion.ac.il/~meshulam

Alex Samorodnitsky
professor, School of Engineering and Computer Science
The Hebrew University of Jerusalem, Jerusalem, Israel
salex@cs.huji.ac.il
http://www.cs.huji.ac.il/~salex

## ABOUT THE AUTHORS

SHACHAR LOVETT graduated from the Weizmann Institute of Science in 2010; his advisors were Omer Reingold and Ran Raz. He is interested in the theory of computing, combinatorics, and coding theory, and in particular in the interplay between structure, randomness, and pseudorandomness.

ROY MESHULAM is a Professor of Mathematics at the Technion. He has worked in a number of areas, including applications of harmonic analysis to combinatorics and extremal problems for spaces of matrices. In recent years his main research interests have been applications of algebraic topology to discrete geometry and random simplicial complexes.

ALEX SAMORODNITSKY graduated from the Hebrew University of Jerusalem in 1999; his advisor was Nathan Linial. He is interested in coding theory, combinatorics, and the theory of computing, and especially in applying tools of analysis and geometry to these areas of discrete math.