Electronic Colloquium on Computational Complexity, Report No. 72 (2008)

Worst Case to Average Case Reductions for Polynomials

Tali Kaufman*
MIT & IAS
kaufmant@mit.edu

Shachar Lovett†
Weizmann Institute of Science
shachar.lovett@weizmann.ac.il

August 10, 2008

Abstract

A degree-d polynomial p in n variables over a field $\mathbb{F}$ is *equidistributed* if it takes on each of its $|\mathbb{F}|$ values close to equally often, and *biased* otherwise. We say that p has *low rank* if it can be expressed as a function of a small number of lower degree polynomials. Green and Tao [GT07] have shown that over large fields (i.e when $d < |\mathbb{F}|$) a biased polynomial must have low rank. They have also conjectured that bias implies low rank over general fields, but their proof technique fails to show that. In this work we affirmatively answer their conjecture. Using this result we obtain a general worst case to average case reductions for polynomials. That is, we show that a polynomial that can be *approximated* by a few polynomials of bounded degree (i.e. a polynomial with non negligible correlation with a function of few bounded degree polynomials), can be *computed* by a few polynomials of bounded degree. We derive some relations between our results to the construction of pseudorandom generators. Our work provides another evidence to the structure vs. randomness dichotomy.

1 Introduction

Let $\mathbb{F}$ be a prime finite field. Let $p : \mathbb{F}^n \to \mathbb{F}$ be a polynomial in n variables over $\mathbb{F}$ of degree at most d. We say that p is *equidistributed* if it takes on each of its $|\mathbb{F}|$ values close to equally often, and *biased* otherwise. We say that p has a *low rank* if it can be expressed as a bounded combination of polynomials of lower degree, and *high rank* otherwise. More formally we consider the following definitions.

Definition 1 (bias)**.** *The* bias *of a function $f : \mathbb{F}^n \to \mathbb{F}$ is defined to be*

$$bias(f) = \mathbb{E}_{X \in \mathbb{F}^n}[\omega^{f(X)}]$$

where ω stands for the $|\mathbb{F}|$ root of unity, i.e. $\omega = e^{\frac{2\pi i}{|\mathbb{F}|}}$.

We use the bias of f as a measure for the distance from uniformity of $f(X) \in \mathbb{F}$ when $X \in \mathbb{F}^n$ is chosen uniformly. The following simple facts explain why we can do so.

Fact 1. *Let $X \in \mathbb{F}^n$ be chosen uniformly. Then:*

*Research supported in part by NSF Awards CCF-0514167 and NSF-0729011.
†Research supported by the Israel Science Foundation (grant 1300/05)

ISSN 1433-8092

- *If $f(X) \in \mathbb{F}$ is uniform then $bias(f) = 0$*
- *If $bias(f) \geq \delta > 0$ then the statistical distance between $f(X)$ and the uniform distribution over $\mathbb{F}$ is at least δ.*
- *If the statistical distance between $f(X)$ and the uniform distribution over $\mathbb{F}$ is δ, then there is some $c \in \mathbb{F}$, $c \neq 0$ s.t. $bias(cf) \geq \delta'$ for $\delta' = \delta/\sqrt{|\mathbb{F}| - 1}$*

Definition 2 (rank)**.** *Let $p(X)$ be a degree d polynomial over $\mathbb{F}^n$. $rank_{d-1}(P)$ is the smallest integer k such that there exist degree $d-1$ polynomials $q_1(X), ..., q_k(x)$, and a function $F : \mathbb{F}^k \to \mathbb{F}$, s.t. $p(X) = F(q_1(X), ..., q_k(X))$.*

Green and Tao [GT07] have shown that over large fields bias implies low rank.

Theorem 2 (Theorem 1.7 in [GT07])**.** *Let $p(X)$ be a degree d polynomial over $\mathbb{F}^n$, where $d < |\mathbb{F}|$. If $bias(p) \geq \delta > 0$, then $rank_{d-1}(p) \leq c(\mathbb{F}, d, \delta)$.*

In their paper, Green and Tao conjecture that the restriction $d < |\mathbb{F}|$ can be removed, but their proof technique breaks down when $d \geq |\mathbb{F}|$. Note that over large fields things might behave differently than over small fields. One important example is the *Inverse Conjecture for the Gowers Norm*. This conjecture roughly says that if the d-derivative of a polynomial is biased then that polynomial has a non-negligible correlation with some polynomial of degree $d - 1$. The *Inverse Conjecture for the Gowers Norm* was proven to be true over large fields by [GT07], but was proven to be false over small fields [GT07, LMS]. One of the main tools used for proving the conjecture over large fields was Theorem 2, that was proven over large fields.

One could ask what is the case with the above theorem, whether it remains true over smaller fields or it becomes false there. We show that the [GT07] result is true over general fields. In this respect, as opposed to the *Inverse Conjecture for the Gowers Norm* case, large and small fields behave similarly.

1.1 Our Main Results

Our first main theorem is a worst case to average case reduction for polynomials. It says that a polynomial that can be approximated by few polynomials of bounded degree, can be computed by few polynomials of bounded degree. We now move to define this rigorously.

Definition 3 (δ-approximation)**.** *We say a function $f : \mathbb{F}^n \to \mathbb{F}$ δ-approximates $p(X)$ if:*

$$|\mathbb{E}_{X \in \mathbb{F}^n}[\omega^{p(X)-f(X)}]| \geq \delta$$

Theorem 3 (Worst-case to average case reduction for polynomials of bounded degree)**.** *Let $p(X)$ be a polynomial of degree d, $g_1, ..., g_c$ polynomials of degree k (where d, c, k are constants) and $F : \mathbb{F}^c \to \mathbb{F}$ a function s.t. the composition $G(x) = F(g_1(X), ..., g_c(X))$ δ-approximates p. Then there exist c' polynomials $h_1, ..., h_{c'}$ and a function $F' : \mathbb{F}^{c'} \to \mathbb{F}$ s.t.*

$$F'(h_1(X), ..., h_{c'}(X)) \equiv p(X)$$

Moreover, $c' = c'(\mathbb{F}, d, c, k, \delta)$ (i.e. independent of n) and each h_i is of the form $p(X + a) - p(X)$ or $g_j(X + a)$ for $a \in \mathbb{F}^n$. In particular, if $k \leq d - 1$ then also $\deg(h_i) \leq d - 1$.

Our first main theorem is obtained as a corollary from our second main theorem, Theorem 4. This theorem shows that bias implies low rank over *general fields*.

Theorem 4 (Bias implies low rank for general fields)**.** *Let $p(X)$ be a degree d polynomial over $\mathbb{F}^n$, s.t. $bias(p) \geq \delta > 0$. Then $rank_{d-1}(p) \leq c(\mathbb{F}, d, \delta)$. That is, there exist degree-$(d-1)$ polynomials $q_1(X), ..., q_c(x)$, and a function $F : \mathbb{F}^c \to \mathbb{F}$, s.t. $p(X) = F(q_1(X), ..., q_c(X))$, and $c = c(\mathbb{F}, d, \delta)$. Moreover, $q_1, ..., q_c$ are derivatives of the form $p(X + a) - p(X)$ where $a \in \mathbb{F}^n$.*

Most of the technical part of the paper is dedicated to proving Theorem 4. The proof is by induction on the degree d of $p(X)$. Notice that for $d = 1$ it holds trivially. So, we assume Theorem 4 to hold for all degrees smaller than d, and prove it for degree d.

2 Significance of Results

Worst case to average case reductions for polynomials. Our first main theorem (Theorem 3) shows that every polynomial, not necessarily biased, that is approximated by few other bounded degree polynomials, can be computed by few bounded degree polynomials. We view this result as a worst case to average case reduction for polynomials. I.e. in order to show that a polynomial can not be approximated by few bounded degree polynomials, it would be sufficient to show that the polynomial can not be computed by few bounded degree polynomials. That later task might be easier. An example when such a scenario is relevant is the following. The papers [GT07, LMS] that disprove the *Inverse Conjecture for the Gowers Norm* needed to show that the symmetric polynomial S_4 over $\mathbb{F}_2$, i.e. $S_4(x_1, ..., x_n) = \sum_{i<j<k<l} x_i x_j x_k x_l$ cannot be approximated by a degree 3 polynomial. Given the current result it could be sufficient (and maybe easier?) to show that S_4 can not be computed by a constant number of degree 3 polynomials.

Proof of the Green Tao Conjecture. Our second main theorem (Theorem 4) shows that over general fields there is a phenomena that bias implies low rank. Green and Tao [GT07] proved this for large fields. They conjectured it to hold also over small fields. We answer their conjecture affirmatively, by showing that the "bias imply low rank" phenomena is robust and holds for all fields.

On the power of induction and relation to pseudorandom generators. Pseudorandom generator for polynomials of degree-d is an efficient procedure that stretches s field elements into $n \gg s$ field elements that can fool any polynomial of degree d in n variables. Pseudorandom generators are mostly interesting over small fields. One can use our second main theorem to provide an alternative proof to the correctness of the pseudorandom generators of [BV] that fools degree d polynomials. Specifically, the generator of [BV] is a XOR of d copies of the generator of Naor and Naor that fools linear functions. The proof of correctness of the [BV] generator of [V] is by induction. The proof assumes the existence of a pseudorandom generator that fools degree $d-1$ polynomial and constructs from it pseudorandom generator that fools degree d polynomial. The proof of the induction step is based on the following. Either the polynomial is unbiased, and hence the generator could fool it. Alternatively, it is biased, and hence again [V] shows that it can be fooled. By our result here, if the polynomial is biased then it has low rank. One can use the property that a generator that can fool a function in the class can fool any composition of few functions from the class to complete the induction step. This proof method is inspired by

the original argument of [BV] the relied on the *Inverse Conjecture for the Gowers Norm* which turned out to be false. The proof of correctness of Viola [V] is clearly more direct. However, we still feel that the original proof strategy of [BV] sheds light on the relations between structure and pseudorandomness in the realm of low degree polynomials.

The "bias imply low rank" idea suggests a robust way to construct pseudorandom generators for some complex function classes based on pseudorandom generators for simpler function classes. This would be done in the spirit of the induction above. Either a function is unbiased , in which case it should be easy to claim that it could be fooled based on the induction assumption, or it is a function of few functions of lower complexity. Use now a property that a generator that can fool a function in the class can fool any composition of few functions from the class. Hence, by *induction* we obtain a construction of pseudorandom generator for functions of higher complexity classes (e.g. degree d polynomials) given pseudorandom generators for functions of lower complexity classes (e.g. linear functions).

Extension to tensors Let $L(x,y)$ be a bilinear form over $\mathbb{F}^n$, i.e. a function of the form

$$L(x,y) = x^t A y$$

where $x, y \in \mathbb{F}^n$ and A is a matrix. There is a close connection between the rank of the matrix and the bias of L. Dixon's Theorem ([MS]) tells us that the bias of L (and in fact, all non-zero Fourier coefficients of L) has absolute value $c(\mathbb{F})^{-rank(A+A^t)}$. The theory of higher dimensional multilinear forms, i.e. tensors, is much less understood. In particular, there is no single notion of tensor rank. We prove, as a direct corollary of Theorem 4, that if we define the rank of a tensor as minimal number of lower degree multilinear forms needed to compute it, then bias imply low rank for tensors.

Theorem 5. *Let $L(X_1,...,X_d)$ be a multilinear form of degree d s.t. $bias(L) \geq \delta > 0$. Then, there exist degree-$(d-1)$ multilinear forms $q_1,...,q_c$, each operating on $d-1$ variables out of $X_1,...,X_d$, and a function $F : \mathbb{F}^c \to \mathbb{F}$, s.t.*

$$\begin{aligned} L(X_1,...,X_d) = F(&q_1(X_1,...,X_{t_1-1},X_{t_1+1},...,X_d),\\ &...,\\ &q_c(X_1,...,X_{t_c-1},X_{t_c+1},...,X_d)) \end{aligned}$$

and $c = c(\mathbb{F},d,\delta)$. Moreover, $q_1,...,q_c$ are derivatives of L.

Proof. We use Theorem 4 on L as a degree d-polynomial, and observe that derivatives of L are sums of d degree-$(d-1)$ multilinear forms in $d-1$ variables of $X_1,...,X_d$. □

2.1 Proof Overview

We will prove that if a degree-d multivariate polynomial over a finite field can be approximated by a function of a constant number of lower degree polynomials, then it in fact be exactly computed by a function of a (larger) constant number of lower degree polynomials. Here and in the paper, constant means independent in the number of variables. In fact, we think of the number of variables as going to infinity, where the rest of the parameters (field size, degree, number of approximating polynomials) as constants. We denote by $p(X)$ a multivariate polynomial, where $X = (x_1,...,x_n) \in \mathbb{F}^n$.

First we reduce the problem to showing that if a polynomial $p(X)$ is biased, then it can be computed by a function of constant number of lower degree polynomials. The reduction is straightforward: if $p(X)$ can be approximated by a function $F(g_1(X), ..., g_k(X))$, where $deg(g_i) < deg(p)$ for all i, then there is some linear combination of the g_i's s.t. $p(X) + a_1 g_1(X) + ... + a_k g_k(X)$ is biased, and thus can be computed by a constant number of lower degree polynomials.

We now describe the proof of the main technical part of the paper, that is, if a degree d polynomial $p(X)$ is biased, then it can be calculated by a constant number of degree $d-1$ polynomials (the constant depending only on the field, the degree d, and the bias of p). The proof is by induction on d. We note that the case $d = 1$ is trivial.

Green and Tao prove the same result [GT07], when the degree d is bounded by the field size, $d < |\mathbb{F}|$. The main contribution of this work is extending this proof for all constant degrees. We will follow closely the proof structure of Green and Tao, and we make one significant divergence which allows us to make the result hold for all constant degrees.

The proof starts, as in the case of the work of Green and Tao, with a lemma of Bogdanov and Viola. Bogdanov and Viola [BV] prove that if a degree-d polynomial $p(X)$ has bias, then it can be well approximated by a constant number of lower degree polynomials. Formally, for every constant $\epsilon > 0$, there is a function F_s and degree $d-1$ polynomials $b_1, ..., b_s$ s.t.

$$\mathbb{P}_{X \in \mathbb{F}^N}[p(X) = F_s(b_1(X), ..., b_s(X))] \geq 1 - \epsilon$$

where s depends only on the field $\mathbb{F}$, the degree d and the required approximation error ϵ. Importantly, s doesn't depend on the number of variables. Bogdanov and Viola in fact show an explicit construction of such a function F and polynomials $b_1, ..., b_s$.

The technical heart of this paper, as well as in the work of Green and Tao [GT07], is to show that when the approximation is good enough, it can in fact be made into an exact computation. Note that we can't use the lemma of Bogdanov and Viola directly, since choosing $\epsilon < |\mathbb{F}|^{-N}$ would result in a non-constant s.

Consider the following partition of $\mathbb{F}^n$ given by the joint distribution of the polynomials $(b_1, ..., b_s)$. For every $c = (c_1, ..., c_s) \in \mathbb{F}^s$, define the region

$$R_c = \{x \in \mathbb{F}^n : \forall i \; b_i(x) = c_i\}$$

The function F_s assigns a value to each region. We say that the joint distribution of $(b_1, ..., b_s)$ is *close to uniform*, if all the regions are roughly of the same size. That is, given $\gamma(s) > 0$, for every $c = (c_1, ..., c_s) \in \mathbb{F}^s$,

$$|R_c| = \frac{|\mathbb{F}|^n}{|\mathbb{F}|^s}(1 \pm \gamma(s)).$$

Green and Tao [GT07] show that a set of polynomials $(b_1, ..., b_s)$ that approximates p in the above sense, can be transformed into a larger set of polynomials called a *regular set* $(g_1, ..., g_t)$ that approximates p and such that the joint distribution of $(g_1, ..., g_t)$ is close to uniform, where t depends only on the field $\mathbb{F}$, the degree d and the required approximation error $\gamma(t)$.

Consider now the regions defined by the polynomials $(g_1, ..., g_t)$. Using averaging arguments the polynomial p is almost constant on most regions. We would like to show that in fact p is constant on all regions. We first show that if p is almost constant on a region, it must be constant on all the region. We then extend this to all regions, assuming p is constant on most regions.

In order to show this, we first recall basic facts regarding derivatives. For a variable $Y \in \mathbb{F}^n$, we define the (discrete) derivative of $p(X)$ in direction Y to be $p_Y(X) = p(X + Y) - p(X)$. It is easy

to see that the degree of X strictly reduces when taking derivatives. We define inductively taking multiple derivatives. For $Y_1, ..., Y_{d+1} \in \mathbb{F}^n$, consider the derivative of $p(X)$ in directions $Y_1, ..., Y_{d+1}$:

$$p_{Y_1,...,Y_{d+1}}(X) = \sum_{I \subseteq [d+1]} (-1)^{d+1-|I|} p(X + \sum_{i \in I} Y_i)$$

since p is a degree d polynomial, this derivative is identically zero. This will play an important role in the proof.

Let R_c be some region on which p is almost constant, and fix some $x_0 \in R_c$. Let $F_{|R_c}$ be the value that F assigns to that region. We will show that if $Y_1, ..., Y_{d+1}$ are chosen uniformly and independently, then there is a positive probability that $x_0 + \sum_{i \in I} Y_i \in R_c$ for all $I \subseteq [d+1]$. Moreover, since almost all points in $x' \in R_c$ are "good", i.e. $p(x') = F_{|R_c}$, there is in fact a positive probability that they all fall in the "good" part of R_c, i.e. that $p(x_0 + \sum_{i \in I} Y_i) = F_{|R_c}$ for all $I \neq \phi$. Plugging this into the derivative equation, and using the fact that it is identically zero, will give that also $p(x_0) = F_{|R_c}$. That is, if a region is almost constant, then it must be fully constant.

So, we need to prove that if $Y_1, ..., Y_{d+1}$ are chosen uniformly, there is a positive probability for all $x_0 + \sum_{i \in I} Y_i$ to fall in R_c and in fact to behave like a uniform point in R_c. In order to do so, we need to use the definition of the region R_c.

Consider the joint evaluation of all the polynomials $g_1, ..., g_t$ on all points $(x_0 + \sum_{i \in I} Y_i)$, i.e. the joint distribution in $\mathbb{F}^{(2^{d+1}-1)t}$ of:

$$\left(g_j(x_0 + \sum_{i \in I} Y_i) : j \in [t], I \subseteq [d+1], I \neq \phi \right)$$

where $Y_1, ..., Y_{d+1}$ are uniform and independent in $\mathbb{F}^n$. (Notice we disallow $I = \phi$, because it corresponds to the evaluations $\{g_j(x_0)\}$, which are fixed since they do not depend on any Y_i.)

If this distribution was uniform (over $\mathbb{F}^{(2^{d+1}-1)t}$), or even close enough to uniform, there was a positive probability that for all $j \in [t]$ and $I \subseteq [d+1]$,

$$g_j(x_0 + \sum_{i \in I} Y_i) = g_j(x_0)$$

Hence, all points $x_0 + \sum_{i \in I} Y_i$ would belong to R_c as required.

However, there is no reason why the joint distribution of $\{g_j(x_0 + \sum_{i \in I} Y_i)\}$ should be close to uniform. One obvious reason is that each polynomial g_j is itself a low degree polynomial, of degree at most $d-1$. Thus, for any $K \subseteq [d+1]$ s.t. $|K| > deg(g_j)$, deriving g_j in directions $\{Y_k : k \in K\}$ yields the zero polynomial, and thus we have the following linear relation:

$$\sum_{I \subseteq K} (-1)^{|K|-|I|} g_j(x_0 + \sum_{i \in I} Y_i) \equiv 0$$

Another reason for correlation is that different polynomials among $g_1, ..., g_t$ can be correlative. For example, we could have that $g_5 = g_1 g_2 + g_3 g_4$.

Green and Tao solve this problem by showing that if there are correlations between the polynomials, apart from the aforementioned linear relations, then using interpolation over $\mathbb{F}$ there must exist a linear functional over $a_1 g_1(X) + ... + a_t g_t(X)$ which is biased. This contradicts the fact, achieved in the construction of the g_i's, that the joint distribution of $(g_1(X), ..., g_t(X) : X \in \mathbb{F}^n)$

is extremely close to uniform. They then show that the linear relations can in fact be dealt with. However, their use of interpolation requires that $d < |\mathbb{F}|$.

We solve the problem in a different way, which allows us to make the result hold for all constant degrees. We transform our original set of polynomials $b_1, ..., b_s$ into a *strongly-regular* set of low degree polynomial $h_1, ..., h_t$, in which we can control all the correlations without using interpolation. The basic idea is that every h_j has an effective degree $\Delta(h_j) \leq deg(h_j)$, s.t. in the set

$$\{h_j(X + \sum_{i \in I} Y_i) : \ j \in [t], \ I \subseteq [d+1], \ |I| \leq \Delta(h_j)\}$$

there are no significant correlations, and any $h_k(X + \sum_{i \in K} Y_i)$ for $|K| > \Delta(h_k)$ can be calculated by a function of $\{h_j(X + \sum_{i \in I} Y_i) : \ j \in [t], \ I \subseteq K, \ |I| \leq \Delta(h_j)\}$.

This definition in fact allows us to prove several results showing that certain sets of evaluations are close to uniform, which are required for the proof.

2.2 Organization

The rest of the paper is organized as follows. We define required notation in Section 3. We define and analyze regularity and strongly regularity of polynomials in Section 4. We prove Theorem 3 and Theorem 4 in Section 5.

3 Preliminaries

$\mathbb{F}$ if a fixed prime field. We work with constant degree polynomials over $\mathbb{F}^n$. We denote by capital letters $X, Y, ...$ variables in $\mathbb{F}^n$, and by small letters $x, y, a, ...$ values in $\mathbb{F}^n$. We use the notation $\mathbb{P}$ for probability measure. Degree of a polynomial will always mean total degree. Unless otherwise specified, when we speak of a degree d polynomial, we mean in fact a polynomial of total degree at most d. For a set of variables $Y_1, Y_2, ... \in \mathbb{F}^n$ we denote by $Y_I = \sum_{i \in I} Y_i$, and similarly for a set of values $y_1, y_2, ... \in \mathbb{F}^n$. We write $u = v(1 \pm \epsilon)$ for $u \in [v(1-\epsilon), v(1+\epsilon)]$. When we speak of a *growth function*, we mean any monotone function $\mathcal{F} : \mathbb{N} \to \mathbb{N}$ (for example, $\mathcal{F}(n) = 2^{n^2}$). We shorthand the set $\{1, 2, ..., t\}$ by $[t]$.

Definition 4 (close to uniform)**.** *The joint distribution of the polynomials* $(g_1, ..., g_s)$ *is* γ-close to uniform/almost independent*, for* $\gamma = \gamma(s) > 0$*, if for every* $(c_1, ..., c_s) \in \mathbb{F}^s$*,*

$$\mathbb{P}_{X \in \mathbb{F}^n}(\forall i \in [c], g_i(X) = c_i) = (1 \pm \gamma(s)) \frac{|\mathbb{F}|^s}{|\mathbb{F}|^n}.$$

4 Regularity of polynomials

As we discussed in the introduction, the notion of regularity plays a major role in our proof. Green and Tao in [GT07] suggested one notion of regularity (we refer to it henceforth as *regularity*) which limited their proof to work only for large fields (i.e. $d < |\mathbb{F}|$). We suggest a stronger notion of regularity (noted henceforth as *strong regularity*). This new notion of strong regularity is essential for obtaining a result for general fields. In the following we review the regularity definitions given by Green and Tao. Then, we present the notion of strong regularity and show that every set of polynomials which approximates a polynomial p can be transformed into a larger set that

approximates p and is also strongly regular. We end this section by showing that strong regularity implies almost independence for sets of variables that forms some specific structures. This almost independence is the crux of the proof of Theorem 4.

Definition 5 (Regularity of polynomials)**.** *Let $\mathcal{F}$ be any growth function. A set of polynomials $\{g_1, ..., g_m\}$ is called $\mathcal{F}$-regular if any linear combination $\alpha_1 g_1(X) + ...\alpha_m g_m(X)$ cannot be expressed as a function of at most $\mathcal{F}(m)$ polynomials of degree $k-1$, where $k = \max\{\deg(g_i) : \alpha_i \neq 0\}$ (i.e. k is the maximal degree of g_i appearing in the linear combination).*

Notice we use a growth function $\mathcal{F}(m)$ instead of a specific number. The reason is that in the application we would not be able to control the number m, and would only care about the relation between the number of polynomials (m) and the strength of the regularity of the set ($\mathcal{F}(m)$).

Green and Tao also define the notion of a refinement of a set of polynomials. Informally, a set $\{g_1, ..., g_m\}$ is a refinement of $\{f_1, ..., f_s\}$ if for any $i \in [s]$, $f_i(x)$ can be computed given the values of $\{g_1(x), ..., g_m(x)\}$.

Definition 6 (Refinement)**.** *A set of polynomials $\{g_1, ..., g_m\}$ is a refinement of $\{f_1, ..., f_s\}$ if for any $i \in [s]$ there exists a function $F_i : \mathbb{F}^m \to \mathbb{F}$ s.t.*

$$f_i(X) = F_i(g_1(X), ..., g_m(X))$$

Green and Tao prove that for any growth function $\mathcal{F}$, any set of polynomials $F = \{f_1, ..., f_s\}$ can be refined to a $\mathcal{F}$-regular set $\{g_1, .., g_m\}$, s.t. m depends only on s, $\mathcal{F}$ and the maximal degree in F. Importantly, m is independent of n.

We now discuss the way Green and Tao use the regularity condition, and why it fails to work when $d > |\mathbb{F}|$. We will then introduce our definition for strong regularity, which overcomes this obstacle.

As we discussed in the proof overview, if $\{g_1, ..., g_m\}$ are $\mathcal{F}$-regular for a large enough $\mathcal{F}$, then the joint distribution of

$$\{g_1(X), ..., g_m(X) : X \in \mathbb{F}^n\}$$

is close to uniform. Green and Tao need in fact a strong condition from the polynomials $g_1, ..., g_m$ in the process of their proof. Let $Y_1, ..., Y_{d+1} \in \mathbb{F}^n$ be new independent chunks of variables. They require that for any $x_0 \in \mathbb{F}^n$,the joint distribution of

$$\{g_i(x_0 + \sum_{i \in I} Y_i) : |I| \leq deg(g_i)\}$$

is also close to uniform. They prove this is true if the field is large ($|\mathbb{F}| > d$). However, over small fields, this doesn't hold in general, as the following example shows.

Example 6. *Consider the symmetric polynomial S_4 over $\mathbb{F}_2$, i.e.*

$$S_4(x_1, ..., x_n) = \sum_{i<j<k<l} x_i x_j x_k x_l$$

Consider the fourth derivative of S_4, i.e. the polynomial in $X, Y_1, ..., Y_4$

$$G(X, Y_1, ..., Y_4) = \sum_{I \subseteq [4]} S_4(X + \sum_{i \in I} Y_i)$$

This polynomial corresponds to the 4-th Gowers Norm of S_4, and as was shown in [GT07] and [LMS], it has bias $1/8$. Thus, the joint distribution of the set

$$\{S_4(x_0 + \sum_{i \in I} Y_i) : |I| \leq \deg(S_4)\}$$

is not close to uniform. This stands in contrast to the fact that $S_4(X)$ is equidistributed over $\mathbb{F}_2$.

Our definition for *strong-regularity* avoids this obstacles by allowing to effectively reduce the degree of a polynomial, if it's high-order derivatives can be calculated from lower-order ones. In fact, for any polynomial g_i we declare an effective degree $\Delta(g_i) \leq \deg(g_i)$. We require that the set

$$\{g_i(X + \sum_{i \in I} Y_i) : \ i \in [m], |I| \leq \Delta(g_i)\}$$

is almost uniform, while for every g_k and K s.t. $|K| > \Delta(g_k)$, $g_k(X + \sum_{iinK} Y_i)$ can be calculated by a function of $\{g_i(X + \sum_{i \in I} Y_i) : \ i \in [m], I \subseteq K, |I| \leq \Delta(g_i)\}$

We now move to formally define our notion of strong regularity, and to show it implies the almost independence/total dependence structure we have just described. We first define the notion of a derivative space.

Definition 7 (Derivative space)**.** *For a set of polynomials $F = \{f_1(X), ..., f_s(X)\}$ we define:*

$$Der(F) = \{f_i(X + a) - f_i(X) : i \in [s], \ a \in \mathbb{F}^n\}$$

Similarly, for a set of polynomials in several variable chunks $F = \{f_1(Y_1, ..., Y_k), ..., f_s(Y_1, ..., Y_k)\}$ $(Y_1, ..., Y_k \in \mathbb{F}^n)$ we define:

$$Der(F) = \{f_i(Y_1 + a_1, ..., Y_k + a_k) - f_i(Y_1, ..., Y_k) : \\ i \in [s], \ a_1, ..., a_k \in \mathbb{F}^n\}$$

Notice that if the maximal degree of polynomials in F is k, then the maximal degree of polynomials in $Der(F)$ is at most $k - 1$. We now formally define strong regularity. We recall that for a set of variables $Y_1, Y_2, ...$, we shorthand $Y_I = \sum_{i \in I} Y_i$.

Definition 8 (Strong regularity of polynomials)**.** *Let $\mathcal{F}$ be any growth function. Let $G = \{g_1, ..., g_m\}$ be a set of polynomials and $\Delta : G \to \mathbb{N}$ be a mapping from G to the natural numbers. We say the set G is strongly $\mathcal{F}$-regular with effective degree Δ if:*

1. *For any $i \in [m]$, $1 \leq \Delta(g_i) \leq \deg(g_i)$.*
2. *For any $i \in [m]$ and $r > \Delta(g_i)$, let X and $Y_1, Y_2, ..., Y_r$ be variables in $\mathbb{F}^n$. There exist a function $F_{i,r}$ s.t.*

$$g_i(X + Y_{[r]}) = \\ F_{i,r}\left(g_j(X + Y_J) : j \in [m], \ J \subseteq [r], \ |J| \leq \Delta(g_j)\right)$$

3. *For any $r \geq 0$, let X and $Y_1, ..., Y_r$ be variables in $\mathbb{F}^n$. Let $\{\alpha_{i,I}\}_{i\in[m],I\subseteq[r],|I|\leq\Delta(g_i)}$ be any collection of field elements, not all zero. Let $a(X, Y_1, ..., Y_r)$ stand for the linear combination:*

$$a(X, Y_1, ..., Y_r) = \sum_{i\in[m],I\subseteq[r],|I|\leq\Delta(g_i)} \alpha_{i,I} g_i(X + Y_I)$$

Let $G' \subseteq G$ be the set of all g_i's which appear in a, i.e.:

$$G' = \{g_i \in G : \exists I \ \alpha_{i,I} \neq 0\}$$

There <u>does not</u> exist polynomials $h_1, ..., h_l \in Der(G')$, $l \leq \mathcal{F}(m)$ s.t. $a(X, Y_1, ..., Y_r)$ can be expressed as:

$$H(h_1(X + Y_{I_1}), ..., h_l(X + Y_{I_l}))$$

for $I_1, ..., I_l \subseteq [r]$ and some function $H : \mathbb{F}^l \to \mathbb{F}$.

If the set G satisfies only (1) *and* (2)*, we say G is* pre-strong-regular *(notice that $\mathcal{F}$ appears only in* (3)*).*

We first prove, similar to the proof in [GT07], that any set of polynomials can be refined to a strong $\mathcal{F}$-regular set, where the size of the resulting set depends only on the size of the original set, and the maximal degree of polynomials in it. Also, the refining set is contained in the space of iterated derivatives of the original polynomials.

We now formally define the space of iterated derivatives.

Definition 9 (Space of iterated derivatives)**.** *For a polynomial set F, we define its iterated derivative set Der_C to be the set of taking at most C derivatives of F, i.e.*

$$Der_0(F) = F$$

$$Der_C(F) = Der(Der_{C-1}(F)) \cup Der_{C-1}(F)$$

Lemma 7 (Strong-Regularity Lemma)**.** *Let $\mathcal{F}$ be any growth function. Let $F = \{f_1, ..., f_s\}$ be a set of polynomials of maximal degree k. There exist a refinement $G = \{g_1, ..., g_m\}$ of F s.t.*

1. *The maximal degree of polynomials in G is also at most k*
2. *The set G is strong $\mathcal{F}$-regular.*
3. *The size m of G is a function of only $\mathcal{F}$, s and k. Importantly, it is independent of n.*
4. *There exists $C = C(\mathcal{F}, s, k)$ s.t. $G \subseteq Der_C(F)$*

Proof. We will start by defining a pre-strong-regular set G from F, and will keep refining it until we reach a strong $\mathcal{F}$-regular set. Our set G will also be in $Der_i(F)$ at the i-th iteration. We will finish by showing that the refinement process must end in a finite number of steps.

We start by defining $\Delta : F \to \mathbb{N}$ by $\Delta(f_i) = \deg(f_i)$, and set the initial value of G to be F. To show that the initial G is pre-strong-regular with effective degree Δ, observe that for any

$r > \deg(f_i)$, deriving f_i r-times yields the zero polynomial. Thus, if $Y_1, ..., Y_r$ are variables, we have the identity:

$$f_i(X + Y_{[r]}) = \sum_{I \subsetneq [r]} (-1)^{r-|I|+1} f_i(X + Y_I)$$

Since we can do this for any $r > \deg(f_i)$, we can continue and express $f_i(X + Y_{[r]})$ as a linear combination of $\{f_i(X + Y_I) : I \subseteq [r],\ |I| \leq \deg(f_i)\}$. Thus, G is pre-strong-regular with effective degree Δ.

We will continue to refine G as long as it is not strong $\mathcal{F}$-regular. Assume $G = \{g_1, ..., g_m\}$ at some iteration is not strong-$\mathcal{F}$-regular. By definition, there is some $r \geq 0$ and coefficients $\{\alpha_{i,I}\}_{i \in [m], I \subseteq [r], |I| \leq \Delta(g_i)}$ s.t. the linear combination:

$$a(X, Y_1, ..., Y_r) = \sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(g_i)} \alpha_{i,I} g_i(X + Y_I)$$

can be expressed as a function of $l \leq \mathcal{F}(m)$ polynomials $h_1, ..., h_l \in Der(G')$, where $G' = \{i \in [m] : \exists I\ \alpha_{i,I} \neq 0\}$ is the set of all g_i's participating in the linear combination.

Let g_{i_0} be a polynomial of maximal degree k in G' and let I_0 be a maximal I in respect to inclusion s.t. $\alpha_{i_0,I_0} \neq 0$. Notice that we must have that $|I_0| \leq \Delta(g_{i_0})$. We have:

$$\sum_{i \in [m], I \subseteq [r], |I| \leq \Delta(g_i)} \alpha_{i,I} g_i(X + Y_I) =$$
$$H(h_1(X + Y_{J_1}), ..., h_l(X + Y_{J_l}))$$

for some function $H : \mathbb{F}^l \to \mathbb{F}$.

Notice first that $\deg(h_i) \leq k - 1$ for all $i \in [l]$. Substitute in the expression $Y_i = 0$ for all $i \notin I_0$. We get that $g_{i_0}(X + Y_{I_0})$ can be expressed as a function of $\{g_{i_0}(X + Y_J) : J \subsetneq I_0\}$, $\{g_j(X + Y_J) : j \neq i_0,\ J \subseteq I_0,\ |J| \leq \Delta(g_j)\}$ and $\{h_j(X + Y_J) : J \subseteq I_0,\ |J| \leq \deg(h_j)\}$. Thus, if we add the polynomials $h_1, ..., h_l$ to G (and set $\Delta(h_i) = \deg(h_i)$), we can reduce $\Delta(g_{i_0})$ to $|I_0| - 1$. If we reduced it to zero, we can remove g_{i_0} entirely from G. The resulting G will be our set for the next iteration.

In order to prove that the refinement process ends after a finite number of iterations (depending on the initial size of F and its maximal degree), notice that at each iteration, the sum of $\Delta(g_i)$ for all $g_i \in G$ with some degree d' reduces by at least 1, where the new polynomials added are all of degree strictly smaller than d', and their number is bounded (as a function of $\mathcal{F}$ and the size of G at the beginning of the iteration). So the total number of iterations is some Ackermann-like function of the initial number of polynomials, their maximal degree and the growth function $\mathcal{F}$. □

4.1 Almost independence by strong regularity

We continue by showing that strong regularity induces almost independence/total dependence structure over general sets of variables. The lemmas we derive are the main technical building blocks in the proof of Theorem 4.

We start with a lemma correlating applications of g_i on sums below the effective degree Δ to all sums over a set of variables.

Lemma 8. *Let $G = \{g_1, ..., g_m\}$ be a strong-regular set with effective degree Δ. Let $x, x' \in \mathbb{F}^n$ be two points s.t. $g_i(x) = g_i(x')$ for all $i \in [m]$. Let $y'_1, ..., y'_k \in \mathbb{F}^n$ be values for some $k \geq 1$, and let $Y_1, ..., Y_k \in \mathbb{F}^n$ be k random variables. Then the following two events are equivalent:*

1. *$A = [g_i(x + Y_I) = g_i(x' + y'_I)$ for all $i \in [m]$ and $I \subseteq [k]]$*
2. *$B = [g_i(x + Y_I) = g_i(x' + y'_I)$ for all $i \in [m]$ and $I \subseteq [k]$ s.t. $1 \leq |I| \leq \Delta(g_i)]$*

Proof. It is obvious that if A holds then also B holds. Assume that B holds, i.e. that

$$g_i(x + Y_I) = g_i(x' + y'_i)$$

for all $i \in [m]$ and $I \subseteq [k]$ s.t. $|I| \leq \Delta(g_i)$. Take some I s.t. $I > \Delta(g_i)$. We need to show that also $g_i(x + Y_I) = g_i(x' + y'_I)$. Since $|I| > \Delta(g_i)$ we know by the strong regularity of G that there is a function $F_{i,I}$ s.t.

$$\begin{aligned} &g_i(X + Y_I) = \\ &F_{i,I}\left(g_j(X + Y_J) : \ j \in [m], \ J \subseteq I, \ |J| \leq \Delta(g_j)\right) \end{aligned}$$

By first substituting $X = x$ to compute $g(x + Y_I)$, and then substituting $X = x'$ and $Y_j = y'_j$ to compute $g(x' + y'_I)$, and using that both $g_j(x) = g_j(x')$ for all $j \in [m]$ and the assumption that B holds, we get that also $g_i(x + Y_I) = g_i(x' + y'_I)$. □

Our next lemma shows that certain evaluations of the polynomials $g_1, ..., g_m$ on linear combinations of the inputs are almost independent, assuming the linear combinations don't have too many non-zero entries. Remember that we are in the process of proving Theorem 4 for degree d by induction. Thus, we assume it to hold for all degrees $d' < d$, and in particular to all linear combinations of $g_1, ..., g_m$.

Lemma 9. *Let $\gamma = \gamma(m)$ be an error term. Let $Y_1, ..., Y_k \in \mathbb{F}^n$ be random variables for some $k \geq 1$. Assume $\mathcal{F}$ is large enough (as a function of γ and k). Assume $g_1, ..., g_m$ are strong $\mathcal{F}$-regular with effective degree Δ. For any non-empty $I \subseteq [k]$ let $x_I \in \mathbb{F}^n$ be some point, and $a^{(I)} = (a_1^{(I)}, ..., a_k^{(I)}) \in \mathbb{F}^k$ s.t.*

- *$a_i^{(I)} \neq 0$ for all $i \in I$*
- *$a_i^{(I)} = 0$ for all $i \notin I$*

Then the joint distribution of

$$\left(g_i(x_I + \sum_{i \in I} a_i^{(I)} Y_i) : i \in [m], I \subseteq [k], 1 \leq |I| \leq \Delta(g_i) \right)$$

is γ-close to the uniform distribution on $\mathbb{F}^{\sum_{i=1}^{m} \sum_{j=1}^{\Delta(g_i)} \binom{k}{j}}$.

We need the following simple lemma for the proof of Lemma 9. It states that a random derivative of a biased polynomial is also biased.

Lemma 10. *Let $h(Y_1, ..., Y_k)$ be a polynomial with bias δ. Let h' be the derivation of h in variables $Y_1, ..., Y_r$ along the directions $Z_1, ..., Z_r$, $(r \leq k)$ i.e.*

$$h'(Y_1, ..., Y_k, Z_1, ..., Z_r) = \sum_{w \in \{0,1\}^r} (-1)^{|w|} h(Y_1 + w_1 Z_1, ..., Y_r + w_r Z_r, Y_{r+1}, ..., Y_k)$$

where $|w|$ denotes the Hamming weight of w. Then $bias(h') \geq \delta^{2^r}$.

Proof. We apply Cauchy-Schwarz. It's enough to prove for $k = 2$ and $r = 1$ because we can group variables.

$$\begin{aligned} bias(h') =& \mathbb{E}_{Y_1, Y_2, Z_1 \in \mathbb{F}^n}[\omega^{h(Y_1, Y_2) - h(Y_1 + Z_1, Y_2)}] = \\ & \mathbb{E}_{Y_2 \in \mathbb{F}^n}[\left(\mathbb{E}_{Y_1 \in \mathbb{F}^n}[\omega^{h(Y_1, Y_2)}]\right)^2] \geq \\ & \left(\mathbb{E}_{Y_1, Y_2 \in \mathbb{F}^n}[\omega^{h(Y_1, Y_2)}]\right)^2 = \delta^2 \end{aligned}$$

□

Proof. (of Lemma 9) We start by using the well known fact, that if a distribution over $\mathbb{F}^r$ is not uniform, it must have some biased functional. If the distribution we study is γ-far from uniform, then there must be a linear functional on $\{g_i(x_I + \sum_{i \in I} a_i^{(I)} Y_i) : \ i \in [m], \ I \subseteq [k], \ |I| \leq \Delta(g_i)\}$ with some non-negligible bias depending on γ. We will prove that if we assume that, we reach a contradiction.

Denote by $Y'_I = \sum_{i \in I} a_i^{(I)} Y_i$, and notice it depends on exactly the same set of variables from $Y_1, ..., Y_k$ as Y_I. By our assumption, there exist coefficients $\{\alpha_{i,I}\}$, not all zero, s.t. the polynomial

$$h(Y_1, ..., Y_k) = \sum_{i \in [m], I \subseteq [k], |I| \leq \Delta(g_i)} \alpha_{i,I} g_i(x_I + Y'_I)$$

has bias at least ρ, where ρ is a function of γ, k and m only (and not of n).

Fix I_0 maximal with regards to inclusion s.t. not all α_{i,I_0} are zero. Since we just care about the bias of h under random $Y_1, ..., Y_k$, we can multiply each Y_i by some non-zero coefficient. We thus assume w.l.o.g that $a_i^{(I_0)} = 1$ for all $i \in I_0$. Let $|I_0| = r$. We assume w.l.o.g that $I_0 = \{1, 2, ..., r\}$. Notice that $Y'_{[r]} = Y_{[r]}$. We also shorthand $x = x_{[r]}$.

Let g_{i_0} be a polynomial with maximal degree $d'' \leq d' < d$ s.t. $\alpha_{i_0, I_0} \neq 0$.

We derive now once each of the variables in $Y_1, ..., Y_r$. Let $\{Z_i\}_{i=1..r}$ be new variables in $\mathbb{F}^n$, and consider:

$$h'(Y_1, ..., Y_k, Z_1, ..., Z_r) = \sum_{w \in \{0,1\}^r} (-1)^{|w|} h(Y_1 + w_1 Z_1, ..., Y_r + w_r Z_r, Y_{r+1}, ..., Y_k)$$

First, by Lemma 10, h' has bias at least $\rho' = \rho^{2^k}$.

Now, consider what happens to a term $g_i(x + Y'_I)$ in h after the derivation. If $I \neq [r]$, by the maximality of I_0 there must exist $i' \in [r]$ s.t. $i' \notin I$. Thus, deriving $Y_{i'}$ zeroes out $g_i(x + Y'_I)$.

So, the only terms remaining in h' come from terms in h of the form $g_i(x+Y_{[r]})$. Thus, h' does not depend on Y_i for $i \notin [r]$, and also all the g_i's remaining must have $\Delta(g_i) \geq r$ (because $g_i(x+Y_{[r]})$ appeared in h with non-zero coefficient). Thus we can write:

$$\begin{aligned} h' =& h'(Y_1,...,Y_r,Z_1,...,Z_r) = \\ & \sum_{i\in[m]} \alpha_{i,[r]} \sum_{w\subseteq[r]} (-1)^{|w|} g_i(x+Y_{[r]}+Z_w) \end{aligned}$$

We now make an important observation. Notice that h' depends only on the sum $Y_{[r]}$, and not on the individual $Y_1,...,Y_r$. So we can substitute $W = x+Y_{[r]}$ and get:

$$\begin{aligned} h' =& h'(W,Z_1,...,Z_r) = \\ & \sum_{i\in[m]} \alpha_{i,[r]} \sum_{w\subseteq[r]} (-1)^{|w|} g_i(W+Z_w) \end{aligned}$$

We have assumed that G is strong $\mathcal{F}$-regular. We will show now that if we choose $\mathcal{F}$ large enough, we have already reached a contradiction. Notice the polynomials $g_i(W+Z_w)$ are exactly those which appear in the regularity requirements (where X is replaced here by W, and $Y_1, Y_2, ...$ by $Z_1, Z_2, ...$). Let G' denote the set of g_i's s.t. g_i appear in h' with non-zero coefficient.

We assume by induction that Theorem 4 holds for $d'' < d$ and for all n. Since all polynomials $g_i \in G$ have degree at most $d-1$, then also $\deg(h') \leq d-1$, and so we can apply Theorem 4 on h'. So, since h' has bias ρ', there must exist polynomials $q_1,...,q_t \in Der(h')$ s.t.

$$\begin{aligned} & h'(W,Z_1,...,Z_r) = \\ & Q(q_1(W,Z_1,...,Z_r),...,q_t(W,Z_1,...,Z_r)) \end{aligned}$$

for some function $Q: \mathbb{F}^t \to \mathbb{F}$, s.t. $t = t(\rho', d'')$. Moreover, since every polynomial q_i is of the form $h'(W+a_0, Z_1+a_1, ..., Z_r+a_r) - h'(W,Z_1,...,Z_r)$ for some constants $a_0,...,a_r \in \mathbb{F}^n$, and h' is the sum of $g_i(W+Z_w)$, we can decompose each q_i to a sum of at most 2^r polynomials of the form $g_i(W+Z_w+a) - g_i(W+Z_w) \in Der(G')$ for $w \subseteq \{0,1\}^r$. Let $q'_1,...,q'_{t'} \in Der(G')$ denote these decomposed polynomials. We thus have that:

$$h'(W,Z_1,...,Z_r) = Q'(q'_1(W+Z_{I'_1}),...,q'_{t'}(W+Z_{I'_{t'}}))$$

for some function $Q': \mathbb{F}^{t'} \to \mathbb{F}$, $t' = 2^r t$ and $I'_1,...,I'_t \subseteq [r]$. We got that we can compute

$$h'(W,Z_1,...,Z_r) = \sum_{i\in[m]} \alpha_{i,[r]} \sum_{w\subseteq[r]} (-1)^{|w|} g_i(W+Z_w)$$

as a function of t' polynomials of degree strictly smaller than d''. If we have $\mathcal{F}(m) > t'$ this is a contradiction to the strong $\mathcal{F}$-regularity of $g_1,...,g_m$.

Summarizing, there can be no linear combination of $\{g_i(x+Y_I) : I \in S, 1 \leq |I| \leq \Delta(g_i)\}$ which has bias more than ρ, and so the distribution is γ-close to uniform. □

A Useful corollary of Lemma 9 and Lemma 8 is the following.

Corollary 11. *Let $x, x' \in \mathbb{F}^n$ be two points s.t. $g_i(x) = g_i(x')$ for all $i \in [m]$. Let $y'_1,...,y'_k \in \mathbb{F}^n$ be values for some $k \geq 1$, and let $Y_1,...,Y_k \in \mathbb{F}^n$ be k random variables. Then*

$$\begin{aligned} & \mathbb{P}\left[g_i(x+Y_I) = g_i(x'+y'_I) \ \forall\ i \in [m],\ I \subseteq [k]\right] = \\ & |\mathbb{F}|^{-\sum_{i=1}^{m}\sum_{j=1}^{\Delta(g_i)}\binom{k}{j}}(1\pm\gamma) \end{aligned}$$

5 From approximation to computation: Proof of Theorems 3 and 4

In this section we prove Theorem 3 and Theorem 4. We start with the proof of Theorem 3 which follows directly from Theorem 4. Assume $F(g_1(X),...,g_c(X))$ δ-approximates $p(X)$. Develop $\omega^{F(z_1,...,z_c)} : \mathbb{F}^c \to \mathbb{C}$ in the Fourier basis. If $F(g_1(X),...,g_c(X))$ δ-approximates $p(X)$, there must exist some Fourier coefficient which δ'-approximates p , where $\delta' \geq \delta|\mathbb{F}|^{-c}$. That means, there exist $\alpha_1,...,\alpha_c \in \mathbb{F}$ s.t. the polynomial

$$p'(x) = p(x) - (\alpha_1 g_1(x) + ... + \alpha_c g_c(x))$$

has bias at least δ'. Using Theorem 4 we get that there must exist at most c' derivatives of p' which computes it exactly. We can now use them and $\alpha_1 g_1 + ...\alpha_c g_c$ to compute p.

In the remaining of this section we prove Theorem 4. The proof will be by induction on the degree d of the polynomial (notice that for $d = 1$ Theorem 4 is trivial). Let $p(X)$ stand for a degree d polynomial with bias δ. The proof starts by a lemma of Bogdanov and Viola [BV], showing that if a polynomial is biased, then it can be well approximated by a function a small number of degree $d-1$ polynomials. This was also the starting point in the work of Green and Tao:

Lemma 12 (Bias imply approximation by few lower degree polynomials)**.** *Let $p(X)$ be a polynomial of degree d with bias δ. For any $\epsilon > 0$ there exist polynomials $f_1(X),...,f_s(X)$ of degree at most $d-1$ and a function $F : \mathbb{F}^s \to \mathbb{F}$ s.t.*

$$\mathbb{P}_{X\in\mathbb{F}^n}[F(f_1(X),...,f_s(X)) \neq p(X)] < \epsilon$$

The number s of the polynomials depends only on δ and ϵ. Moreover, $f_1,...,f_s \in Der(p)$.

The following lemma is the technical heart of the paper.

Lemma 13 (Approximation by few lower degree polynomials imply computation by few lower degree polynomials)**.** *Let $p(X)$ be a polynomial of degree d, $f_1,...,f_s$ polynomials of degree $d-1$, ($s = O(1)$) and $H : \mathbb{F}^s \to \mathbb{F}$ a function s.t. the composition $H(f_1(X),...,f_s(X))$ ϵ_d-approximates p, where $\epsilon_d = 2^{-\Omega(d)}$ Then there exist s' polynomials $f'_1,...,f'_{s'}$ and a function $H' : \mathbb{F}^{s'} \to \mathbb{F}$ s.t.*

$$H'(f'_1(X),...,f'_{s'}(X)) \equiv p(X)$$

Moreover, $s' = s'(d,s)$ (i.e. independent of n) and each f'_i if of the form $p(X+a) - p(X)$ or $f_j(X+a)$ for $a \in \mathbb{F}^n$.

Thus, to complete the proof of Theorem 4, it remains to prove Lemma 13.

In the following we prove Lemma 13. The main technical tool that we will use are Lemmas 8 and 9. We start the proof of Lemma 13 by refining $F = \{f_1,...,f_s\}$ to a strong-regular set. Let $\mathcal{F}$ be a large enough growth function (to be determined later). By Lemma 7 there exists a set $G = \{g_1,...,g_m\}$ refining F, and an effective degree Δ, s.t. G is strong $\mathcal{F}$-regular with effective degree Δ. Moreover, there exists a $C = C(\mathcal{F},d,\delta)$ s.t. $G \subseteq Der_C(F)$. We know that G also approximates $p(X)$ at least as well as F does. We will prove that it is in fact computes F completely. We can then decompose each $g_i \in Der_C(F)$ as a sum of at most 2^C elements in $Der(p)$ to conclude the result.

Thus, we need to show that G in fact computes $p(X)$ completely. For $c = (c_1,...,c_m) \in \mathbb{F}^m$, denote by $R_c \subseteq \mathbb{F}^n$ the region

$$R_c = \{x \in \mathbb{F}^n : \forall i\ g_i(x) = c_i\}$$

To show that G computes $p(X)$ is equivalent to showing that $p(X)$ is constant on any region R_c. Thus, we turn to study the regions R_c.

We first show (Lemma 14) that all regions R_c have about the same volume, i.e. that they form an almost uniform division of $\mathbb{F}^n$ to $\mathbb{F}^m$ regions. Since G is a strong regular refitment of F that ϵ_d-approximates p we know that also G ϵ_d-approximates p, i.e. there exists some $H' : \mathbb{F}^m \to \mathbb{F}$ s.t.

$$\mathbb{P}_{X\in\mathbb{F}^n}[H'(g_1(X), ..., g_m(X)) \neq p(X)] < \epsilon_d$$

For every region R_c, let η_c be the probability that p is different from G on that region (G is constant on the region).

$$\eta_c = \mathbb{P}_{X\in R_c}[p(X) \neq G|_{R_c}]$$

Since the average of η_c is at most ϵ_d, and all regions are almost uniform (Lemma 14) there can be at most $\sqrt{\epsilon_d}|\mathbb{F}|^m$ regions on which $\eta_c > \sqrt{\epsilon_d}$. We call these the *bad regions*, and we call the rest of the regions *almost good regions*. Next we show (Lemma 15) that the almost good regions are totally good and p is fixed on them. Last, we use the fact that there are only few bad regions and p is fixed on the rest to conclude that p is also fixed on the bad regions (Lemma 17). Thus, $p(X)$ is in fact constant on all regions. To complete the proof of Lemma 13, it remains to prove Lemmas 14, 15 and 17. The following lemma is a direct implication of Corollary 11.

Lemma 14 (Regions are uniform)**.** *Let $\gamma = \gamma(m) > 0$ be a small enough error term. If $\mathcal{F}$ is large enough than $|R_c| = |\mathbb{F}|^{n-m}(1 \pm \gamma)$, for all $c \in \mathbb{F}^m$.*

Proof. Let $c \in \mathbb{F}^m$ and assume first that R_c is not empty, i.e. there exist some x s.t. $g_i(x) = c_i$ for all $i \in [m]$. We apply Corollary 11 with $k = 1$, $x' = x$ and $y_1 = 0$ and get:

$$\mathbb{P}_{Y_1}[g_i(x + Y_1) = g_i(x),\ \forall\ i \in [m]] = |\mathbb{F}|^{-m}(1 \pm \gamma)$$

Substituting $Y = x + Y_1$ proves the result for R_c.

To show that there can be no empty regions, assume otherwise. Thus, there are at most $|\mathbb{F}|^m - 1$ non-empty cells, and each has volume at most $|\mathbb{F}|^{n-m}(1+\gamma)$. Thus $(|\mathbb{F}|^m - 1)|\mathbb{F}|^{n-m}(1+\gamma) \geq |\mathbb{F}|^n$. If $\gamma(m) < |\mathbb{F}|^{-m}$ we get a contradiction. Thus, there are no empty regions, and so all regions have volume $|\mathbb{F}|^{n-m}(1 \pm \gamma)$. □

Lemma 15 (Almost good regions are good)**.** *Let R_c be a region s.t $\mathbb{P}_{X\in R_c}[p(X) = b] > 1 - 2^{-2(d+1)}$, for some constant $b \in \mathbb{F}$. Then $p(X) = b$ for all $X \in R_c$.*

Before proving the lemma we need the following counting lemma on the number of hypercubes and pairs of hypercubes inside a region, similar to one in [GT07]. However, our technique avoids the need of interpolation.

Lemma 16. *Let $\gamma = \gamma(m) > 0$ be small enough error term, and assume $\mathcal{F}$ is large enough. For any point $R = R_c$ and a point $x \in R$ we have:*

1. *Let $Y_1, ..., Y_{d+1}$ be variables in $\mathbb{F}^n$. Then:*

$$\mathbb{P}_{Y_1,...,Y_{d+1}\in\mathbb{F}^n}[x + Y_I \in R,\ \forall I \subseteq [d+1]] =$$
$$|\mathbb{F}|^{-\sum_{i=1}^{m}\sum_{j=1}^{\Delta(g_i)}\binom{d+1}{j}}(1 \pm \gamma)$$

2. *Let $Y_1, ..., Y_{d+1}, Z_1, ..., Z_{d+1}$ be variables in $\mathbb{F}^n$. For any non-empty $I_0 \in [d+1]$:*

$$\mathbb{P}\left[x + Y_I \in R, x + Z_I \in R,\ \forall I \subseteq [d+1] | Y_{I_0} = Z_{I_0}\right] \leq$$
$$|\mathbb{F}|^m \left(|\mathbb{F}|^{-\sum_{i=1}^{m}\sum_{j=1}^{\Delta(g_i)}\binom{d+1}{j}}\right)^2 (1+\gamma)$$

Proof of Lemma 16. In the following we show that the two conditions of the lemma hold.

1. This is a direct application of Corollary 11 for $k = d+1$, $x' = x$ and $y_1, ..., y_k = 0$.

2. Assume w.l.o.g that $I_0 = \{1, 2, ..., s\}$ for $1 \leq s \leq d+1$. We start by making a linear transformation on the coordinates to bring Y_{I_0} and Z_{I_0} to a single variable. Let $Y'_i = Y_i$ for $i \neq s$ and $Y'_s = Y_1 + ... + Y_s$, and similarly define $Z'_1, ..., Z'_{d+1}$. We write Y_I in the basis of $Y'_1, ..., Y'_{d+1}$. Divide $I = I_s \cup I_{\bar{s}}$ where $I_s = I \cap [s]$ and $I_{\bar{s}} = I \setminus I_s$. We have:

   - If $s \notin I$, $Y_I = \sum_{i \in I} Y'_i$
   - If $s \in I$, $Y_I = Y'_s - \sum_{i \in [s] \setminus I_s} Y'_i + \sum_{i \in I_{\bar{s}}} Y'_i$

   Consider for every I the set T_I of indices of Y'_i which appear in the expansion of Y_I. Notice that for any $T \subseteq [d+1]$ there is exactly one I s.t. $T_I = T$. In particular, in order that $g_i(x + Y_I) = g_i(x)$ for all I, we must have in particular that:

   - For any $I \subseteq [d+1]$ s.t. $s \notin I$ and $|I| \leq \Delta(g_i)$,
   $$g_i(x + Y'_I) = g_i(x)$$
   - For any $I \subseteq [d+1]$ s.t. $s \in I$ and $|I| \leq \Delta(g_i)$,
   $$g_i(x + Y'_s - Y'_{I \cap [s-1]} + Y'_{I \cap \{s+1, ..., d+1\}}) = g_i(x)$$

   Similarly for the Z''s, using the fact that the event $Y_{I_0} = Z_{I_0}$ translates to $Z'_s = Y'_s$:

   - For any $I \subseteq [d+1]$ s.t. $s \notin I$ and $|I| \leq \Delta(g_i)$,
   $$g_i(x + Z'_I) = g_i(x)$$
   - For any $I \subseteq [d+1]$ s.t. $s \in I$ and $|I| \leq \Delta(g_i)$,
   $$g_i(x + Y'_s - Z'_{I \cap [s-1]} + Z'_{I \cap \{s+1, ..., d+1\}}) = g_i(x)$$

   The probability of this event is an upper bound on our required probability. Since our variables
   $$Y'_1, ..., Y'_{d+1}, Z'_1, .., Z'_{s-1}, Z'_{s+1}, ..., Z'_{d+1}$$
   are uniform and independent, we can apply Lemma 9 to show that its probability is the required upper bound. The number of subsets of size $j > 1$ in the above events is $\binom{d+1}{j}$ for the event on the Y''s, and also $\binom{d+1}{j}$ for the event on $Z'_1, ..., Z'_{s-1}, Y'_s, Z'_{s+1}, ..., Z'_{d+1}$. For

$j = 1$ however we have intersection (Y'_s is appearing twice), and so the number of events is $2\binom{d+1}{1} - 1$. Thus,by Lemma 9 the probability of the total event is:

$$|\mathbb{F}|^m \left(|\mathbb{F}|^{-\sum_{i=1}^{m} \sum_{j=1}^{\Delta(g_i)} \binom{d+1}{j}} \right)^2 (1 \pm \gamma)$$

which upper bounds the required probability.

□

We now prove Lemma 15 using Lemma 16. Our proof is similar to the one of [GT07].

Proof of Lemma 15. Let $B \subseteq R$ be the set of all "bad" points $x \in R$ on which $p(x) \neq b$. By our assumption, $|B| < 2^{-2(d+1)}|R|$. Assume B is non-empty, and choose some $x \in B$. Let $Y_1, ..., Y_{d+1}$ be random variables in $\mathbb{F}^n$. Fix small enough $\gamma = \gamma(m)$. By Lemma 16 (1),

$$p_R = \mathbb{P}[x + Y_I \in R,\ \forall I \subseteq [d+1]] \geq |\mathbb{F}|^{-\sum_{i=1}^{m} \sum_{j=1}^{\Delta(g_i)} \binom{d+1}{j}} (1 - \gamma)$$

We now wish to bound the event that when all $X + Y_I$ are in R, some $X + Y_I$ is in B, and then union bound over all possible I.

We start by applying Cauchy-Schwarz to transform the problem to counting pairs of hypercubes. Fix some non-empty $I_0 \subseteq [d+1]$, and let

$$p_B = \mathbb{P}[x + Y_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Y_{I_0} \in B] = \sum_{x_0 \in B} \mathbb{P}[x + Y_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Y_{I_0} = x_0]$$

We need to upper bound p_B.

$$p_B^2 = \left(\sum_{x_0 \in B} \mathbb{P}[x + Y_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Y_{I_0} = x_0] \right)^2 \leq |B| \sum_{x_0 \in B} \mathbb{P}[x + Y_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Y_{I_0} = x_0]^2$$

Introducing new variables $Z_1, ..., Z_{d+1}$ in $\mathbb{F}^n$, we have.

$$p_B^2 \leq |B| \mathbb{P}[x + Y_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Z_I \in R\ \forall I \subseteq [d+1]\ \wedge\ x + Y_{I_0} = x + Z_{I_0}]$$

Thus we get that:

$$p_B^2 \leq |B| |\mathbb{F}|^{-n} \mathbb{P}[x + Y_I \in R,\ x + Z_I \in R\ \forall I \subseteq [d+1] | x + Y_{I_0} = x + Z_{I_0}]$$

By claim (2) in Lemma 16 we get that this probability is at most

$$|B||\mathbb{F}|^{m-n}p_R^2(1+\gamma)$$

By Lemma 14, $|R| = |\mathbb{F}|^n \mathbb{P}_{X\in\mathbb{F}^n}[X \in R] = |\mathbb{F}|^{n-m}(1 \pm \gamma)$. Thus, we have that:

$$p_B^2 \leq \frac{|B|}{|R|}p_R^2(1 \pm 2\gamma) \leq 2^{-2(d+1)}p_R^2$$

and thus $\frac{p_B}{p_R} \leq 2^{-(d+1)}(1 \pm 2\gamma)$. We can now union bound over all non-empty $I_0 \subseteq [d+1]$. The probability that there is some I_0 for which $x + Y_{I_0} \in B$ is at most

$$(2^{d+1} - 1)(2^{-(d+1)} + \gamma) < 1$$

for small enough γ.

Thus, there must exist $y_1, ..., y_{d+1} \in \mathbb{F}^n$ s.t.

$$x + y_I \in R \setminus B$$

for all non-empty $I \subseteq [d+1]$. Equivalently, $p(x + y_I) = b$ for all such I's. However, since $p(X)_{y_1,...,y_{d+1}} \equiv 0$,

$$p(x) = \sum_{I\subseteq[d+1],|I|>0} (-1)^{|I|+1}p(x + y_I)$$

and so if all $p(x + y_I) = b$, then also $p(x) = b$, hence $x \notin B$. So we have proved that B is empty, i.e. p is constant on R. □

We finish the proof of Lemma 13 by proving that if $p(X)$ is constant over almost all regions, then it must be constant over any region.

Lemma 17 (If almost all regions are totally good, all are totally good)**.** *Assume that the fraction of regions on which p is constant is at least $1 - 2^{-(d+2)}$. Then p is constant over any region.*

Proof. Let R be any region, and $x, x' \in R$ two points in R. We need to show that $p(x) = p(x')$. Choose $y'_1, ..., y'_{d+1} \in \mathbb{F}^n$ randomly. The probability that $x' + y'_I$ falls in a bad region for any non-empty $I \subseteq [d+1]$ is $2^{-(d+2)}$ (since regions are almost uniform, see Lemma 14). Thus, applying union bound over all non-empty $I \subseteq [d+1]$ we get that $\{x' + y'_I\}$ fall in good regions for all non-empty I with probability at least $1/2$. Fix some $y'_1, ..., y'_{d+1}$ fulfilling this requirement.

Let $Y_1, ..., Y_{d+1} \in \mathbb{F}^n$ be random variables. Since $g_i(x) = g_i(x')$ for all $i \in [m]$ we can apply Corollary 11:

$$\mathbb{P}\left[g_i(x + Y_I) = g_i(x' + y'_I)\ \forall\ i \in [m],\ I \subseteq [d+1]\right] = |\mathbb{F}|^{-\sum_{i=1}^{m}\sum_{j=1}^{\Delta(g_i)}\binom{d+1}{j}}(1 \pm \gamma)$$

In particular, for small enough γ we get that

$$\mathbb{P}\left[g_i(x + Y_I) = g_i(x' + y'_I)\ \forall\ i \in [m],\ I \subseteq [d+1]\right] > 0$$

Let $y_1, ..., y_{d+1}$ be such assignment to $Y_1, ..., Y_{d+1}$. We thus have that for all non-empty $I \subseteq [d+1]$ and for all $i \in [m]$, $g_i(x + y_I) = g_i(x' + y'_I)$. Since the region of $x' + y'_I$ is good for all non-empty I,

we get that for all non-empty $I \subseteq [d+1]$, $p(x + y_I) = p(x' + y'_I)$. We now use the fact that p is a degree d polynomial. If we derive p $d+1$-times in any direction, we will always get zero. We thus have that for $x, y_1, ..., y_{d+1} \in \mathbb{F}^n$: $\sum_{I \subseteq [d+1]} (-1)^{|I|} p(x + y_I) = 0$. Since the same identity is true for $x', y'_1, ..., y'_{d+1}$, we get that $p(x) = p(x')$. □

Acknowledgement We would like to thank Avi Wigderson, Noga Alon and Terrence Tao for helpful discussions. The second author would like to thank his advisor, Omer Reingold, for his help and support.

References

[GT07] B. Green, T.Tao,The distribution of polynomials over finite fields, with applications to the Gowers norms, preprint, 2007.

[BV] A. Bogdanov and E. Viola. Pseudorandom bits for polynomials via the Gowers norm. In *the 48th Annual Symposium on Foundations of Computer Science (FOCS 2007).*

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

[LMS] S. Lovett, R. Meshulam and A. Samorodnitsky, The Inverse Conjecture for the Gowers Norm is False, to appear in *the 40th ACM Symposium on Theory of Computing (STOC 2008).*

[V] E. Viola, The sum of d small-bias generators fools polynomials of degree d. *Proceedings of the 23th IEEE Conference on Computational Complexity (CCC), 2008*, To appear.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de/