48th Annual IEEE Symposium on Foundations of Computer Science

Pseudorandom bits for polynomials

Andrej Bogdanov* Emanuele Viola†

Abstract

We present a new approach to constructing pseudorandom generators that fool low-degree polynomials over finite fields, based on the Gowers norm. Using this approach, we obtain the following main constructions of explicitly computable generators $G : \mathbb{F}^s \to \mathbb{F}^n$ that fool polynomials over a prime field $\mathbb{F}$:

1. *a generator that fools degree-2 (i.e., quadratic) polynomials to within error $1/n$, with seed length $s = O(\log n)$,*
2. *a generator that fools degree-3 (i.e., cubic) polynomials to within error ϵ, with seed length $s = O(\log_{|\mathbb{F}|} n) + f(\epsilon, \mathbb{F})$ where f depends only on ϵ and $\mathbb{F}$ (not on n),*
3. *assuming the "Gowers inverse conjecture," for every d a generator that fools degree-d polynomials to within error ϵ, with seed length $s = O(d \cdot \log_{|\mathbb{F}|} n) + f(d, \epsilon, \mathbb{F})$ where f depends only on d, ϵ, and $\mathbb{F}$ (not on n).*

We stress that the results in (1) and (2) are unconditional, i.e. do not rely on any unproven assumption. Moreover, the results in (3) rely on a special case of the conjecture which may be easier to prove.

Our generator for degree-d polynomials is the component-wise sum of d generators for degree-1 polynomials (on independent seeds).

Prior to our work, generators with logarithmic seed length were only known for degree-1 (i.e., linear) polynomials (Naor and Naor; SIAM J. Comput., 1993). In fact, over small fields such as $\mathbb{F}_2 = \{0, 1\}$, our results constitute the first progress on these problems since the long-standing generator by Luby, Veličković and Wigderson (ISTCS 1993), whose seed length is much bigger: $s = \exp\left(\Omega\left(\sqrt{\log n}\right)\right)$, even for the case of degree-2 polynomials over $\mathbb{F}_2$.

*adib@dimacs.rutgers.edu. DIMACS, Rutgers — the State University of New Jersey, 96 Frelinghuysen Rd, Piscataway, NJ 08854

†viola@ias.edu. Institute for Advanced Study, 1 Einstein Dr, Princeton, NJ 08540. The author is supported by NSF grant CCR-0324906.

1 Introduction

A *pseudorandom generator* $G\colon D^s \to D^n$ for a class of tests $\mathcal{T}$ is an efficient procedure that stretches s input domain[1] elements into $n \gg s$ output elements such that the distribution of the output of the generator *fools* any test $T \in \mathcal{T}$, $T\colon D^n \to D$, in the sense that the statistical distance between $T(X)$ and $T(G(X))$ is small.

Pseudorandom generators are a central object of theoretical computer science that has found a striking variety of applications in complexity theory, algorithm design, and cryptography, and we refer the reader to the excellent book by Goldreich [8] for background.

A fundamental class of tests $\mathcal{T}$ is that of *low-degree polynomials* over a finite field $D = \mathbb{F}$. The special case of linear polynomials over $\mathbb{F}_2 = \{0, 1\}$ was first studied by Naor and Naor [15] who gave a generator with seed length $s = O(\log n)$ (for error $\epsilon = 1/n$), which is optimal up to constant factors (cf. [2]). This generator, also known as small-bias generator, has been one of the most celebrated results in pseudorandomness, with applications ranging from derandomization [15], to PCP's [5], and to lower bounds [4, 21], just to name a few (cf. references in [5]).

Subsequently, Luby, Veličković, and Wigderson (Theorem 2 in [14]; cf. [20]) built a generator that in particular fools constant-degree polynomials over small fields (e.g., $\mathbb{F}_2$).[2] However, the seed length of their generator is much worse that that of Naor and Naor; specifically, it is $s = \exp\left(O\left(\sqrt{\log n}\right)\right)$ (for $\epsilon = 1/n$). (Alternatively, $n = s^{\Omega(\log s)}$ in [14], whereas $n = 2^{\Omega(s)}$ in [15].) Bogdanov [7] also constructed generators, but only over large fields; in particular the field size must be superlogarithmic in n.[3] Over small fields such as $\mathbb{F}_2$, previous to our work there had been no progress on constructing generators for polynomials since the '93 paper [14], even for the case of quadratic polynomials.

[1] The case $D = \{0, 1\}$ is of particular interest, but in this work we will consider both $D = \{0, 1\}$ as well as other domains.

[2] Their generator actually fools a certain class of depth-2 circuits that in particular can implement polynomials whose number of terms is bounded by $n^{O(1)}$, such as constant-degree polynomials.

[3] [7] also gives generators over small fields, but in this case the seed length is worse than what can be obtained from [14].

0272-5428/07 $25.00 © 2007 IEEE
DOI 10.1109/FOCS.2007.42

1.1 Our results

In this work we construct the following generators.

Theorem 1. *Over any prime field $\mathbb{F}$, there exist the following efficiently computable generators $G : \mathbb{F}^s \to \mathbb{F}^n$:*

1. *a generator that fools quadratic ($d = 2$) polynomials with error $1/n$ and seed length $s = O(\log n)$,*
2. *a generator that fools cubic ($d = 3$) polynomials with error ϵ and seed length $s = O(\log_{|\mathbb{F}|} n) + f(\epsilon, \mathbb{F})$, where f depends on ϵ and $\mathbb{F}$ only (not on n). For constant $|\mathbb{F}|$, the seed length is $s = O(\log_{|\mathbb{F}|} n) + \exp(1/\epsilon^{O(1)})$.*

Note that for the case of quadratic polynomials (Item 1 in Theorem 1) and the case of cubic polynomials (Item 2 in Theorem 1) where ϵ and $|\mathbb{F}|$ are constants, we obtain optimal seed length up to constant factors. In fact, the dependence on n is nearly optimal, as we point out towards the end of this section.

As we explain later, our results are based on the "Gowers norm." Under (a special case of) a conjecture known as "Gowers inverse conjecture," we obtain generators for higher degree polynomials.

Theorem 2. *Assume that the "d vs. $d-1$ Gowers inverse conjecture" (Conjecture 22) holds for a field $\mathbb{F}$ and every degree d. Then there exists an efficiently computable generator $G : \mathbb{F}^s \to \mathbb{F}^n$ that fools degree-d polynomials with error ϵ and seed length $O(d \cdot \log_{|\mathbb{F}|} n) + f(d, \epsilon, \mathbb{F})$, for some function f that depends on d, ϵ, and $\mathbb{F}$, but not on n.*

We remark that the d vs. $d-1$ Gowers inverse conjecture may be significantly easier to prove than the general one (cf. discussion after Conjecture 22). We also point out that a strengthening of known "inverse results" [11, 16] would improve the term $\exp\left(1/\epsilon^{O(1)}\right)$ in Item (2) in Theorem 1 to $1/\epsilon^{O(1)}$.

1.2 Techniques

Our generator for degree d polynomials is the component-wise sum of d independent copies of generators for degree-1, i.e. linear, polynomials. The explicitness of our generator immediately follows from known constructions of generators for linear polynomials, such as [15, 2, 1].

Our proof technique is new and is based on the so-called "Gowers norm," which we call "degree norm." This norm was introduced by Gowers [9, 10] and independently by Alon et al. [3], and has found a wide variety of applications, ranging from arithmetic combinatorics [9, 10, 11], property testing [3], PCP's [17, 16], and lower bounds [19, 21]. For simplicity we focus on the case of $\mathbb{F}_2 = \{0, 1\}$; this case shows all our main ideas and avoids technicalities regarding complex numbers. However, we stress that our techniques apply over arbitrary prime fields.

The degree-d norm of a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ is a real number between 0 and 1 that we denote $U_d(f)$. The key idea of this norm is that

$$\boxed{U_d(f) \text{ is an estimate of the maximum correlation of } f \text{ with degree } d-1 \text{ polynomials.}} \quad (\star)$$

The notion of "correlation" in $(\star)$ is standard and simply means how well the function f can be approximated by degree-$(d-1)$ polynomials: It equals the maximum over all degree $d-1$ polynomials q of the quantity

$$\begin{aligned}&\text{Correlation}(f, q)\\ &\quad := \left|\Pr_X[f(X) = q(X)] - \Pr_X[f(X) \neq q(X)]\right|.\end{aligned}$$

In other words, $(\star)$ says that $U_d(f)$ is as close to 1 as f is close to a degree $d-1$ polynomial. In particular, $U_d(f) = 1$ if f has degree $d-1$, while for a random function f we expect $U_d(f)$ to be close to 0. Specifically, it is known that $U_d(f)^{1/2^d}$ always upper bounds the correlation with degree $d-1$ polynomials (cf. [9, 10, 11, 21]); the converse is known to hold for $d = 2$ with polynomial slackness in the parameters (see, e.g., [11, 16]), for $d = 3$ with exponential slackness in the parameters [11, 16], and is conjectured to hold for any fixed d (see [11, Section 13] and [16]). This latter conjecture is usually referred to as the "Gowers inverse conjecture." In this discussion we ignore both the status and the quantitative aspect of $(\star)$ and we proceed with the intuition behind our approach.

We now explain how we establish the correctness of our generator. We take d independent outputs $W_1, \ldots, W_d$ of a linear generator with sufficiently small bias. Our goal is to show that the distribution

$$W := W_1 + W_2 + \cdots + W_d$$

fools any degree-d polynomial p, where the sum denotes bit-wise xor:

$$\textbf{Goal:} \quad \Pr_{X \in \mathbb{F}_2^n}[p(X) = 0] \approx \Pr_W[p(W) = 0]. \quad (1)$$

The main idea of our analysis is a case analysis based on the value $U_d(p)$.

Case $U_d(p)$ small, fooling the Gowers norm: If $U_d(p)$ is small then the *bias* of the polynomial is small, where the bias is simply the average value of the polynomial (over truly random input X): $\text{Bias}(p(X)) := |\Pr_X[p(X) = 0] - \Pr_X[p(X) = 1]|$. This fact immediately follows from $(\star)$: The bias of the polynomial p is simply the correlation of p with the degree-0 constant function

0, and thus it must be small if $U_d(p)$ is small:

$$\mathrm{Bias}(p(X)) \leq U_d(p). \tag{2}$$

Our approach in this case is to show that *Equation (2) stays true even under the pseudorandom distribution W*. Specifically, we show that

$$\mathrm{Bias}(p(W)) \leq U_d(p), \tag{3}$$

and from these two equations (2) and (3) our goal (1) follows easily (both probabilities in (1) are close to $1/2$).

To prove Equation (3) we make two observations. The first is that the degree-d norm of a polynomial p of degree d equals the bias of a *block-linear* polynomial $q_p(y_1, y_2, \ldots, y_d)$, where each y_i is a block of n variables, and by block-linear we mean that for every i the function $q_p(y_1, y_2, \ldots, y_d)$ is a linear function in y_i. The second observation is that the proof of (2) is based on a Cauchy-Schwarz argument which generalizes to any distribution that can be written as the XOR of d independent distributions, such as W. The combination of these two observations enables us to essentially prove that

$$\begin{aligned}\mathrm{Bias}(p(W)) &\leq \mathrm{Bias}(q_p(W_1, W_2, \ldots, W_d)) \\ &\approx \mathrm{Bias}(q_p(Y_1, Y_2, \ldots, Y_d)) = U_d(p),\end{aligned}$$

where the approximation holds because q_p is block-linear and each of the W_i fools linear tests (using a hybrid argument). This proves Equation (3) and concludes the proof in the case of small $U_d(p)$.

Case $U_d(p)$ large. The basic idea in the case that $U_d(p)$ is large is that by ($\star$) the polynomial p is correlated to a degree-$(d-1)$ polynomial, and thus we can argue by induction. More specifically, we will write p as a function of *few* degree-$(d-1)$ polynomials, and then we use the fact – not too hard to show – that a generator that fools degree-$(d-1)$ polynomials also fools any function of few degree-$(d-1)$ polynomials, where the loss in the error naturally depends on the number of degree-$(d-1)$ polynomials. To get the best parameters, we perform a special analysis in the case of $d = 2$.

1. *Subcase $d = 2$, canonical representations of quadratic polynomials:* For quadratic polynomials, we use a structural result from the theory of quadratic forms that shows that any degree-2 polynomial p is equivalent, up to an invertible linear transformation A, to a degree-2 polynomial *where all the t quadratic terms are on disjoint sets of variables*:

   $$(p \circ A)(x) = x_1x_2 + x_3x_4 + \ldots + x_{2t-1}x_{2t} + \ell(x),$$

   where $\ell(x)$ is of degree 1. As it turns out, the degree norm is invariant under invertible linear transformation and shifts, and depends exponentially on the number t of disjoint quadratic terms. This gives

   $$U_2(p) = U_2(p \circ A) = 2^{-2t}.$$

   Since $U_2(p)$ was assumed to be large, t is small. Applying the inverse linear transformation A^{-1}, this means that p can be written as a function of at most $2 \cdot t + 1$ linear functions (specifically, as a sum of products of 2 linear functions, plus a linear function). This gives the desired compact representation of p in terms of linear polynomials, and concludes the proof for $d = 2$.

2. *Subcase $d > 2$, self-correcting polynomials:* For degree $d > 2$ no structural result as the one above is known to our knowledge. Our approach in this case is to use the *self-correcting property* of polynomials. Specifically, from ($\star$) we know that there is a degree $d-1$ polynomial q that correlates well with p. From this we infer that p can be approximated by a function of few degree $d-1$ polynomials up to a small error (the gain is that this latter error is much smaller than the error of the original correlation). For this we use the following *self-correcting* property of low-degree polynomials: The evaluation of the degree-d polynomial p at a given point x can be obtained as the evaluation $p(x+a)$ of the polynomial p at a random shift $x + a$, for which we use q, minus the "derivative" $\mathrm{D}_a p$ of the polynomial p, $\mathrm{D}_a p(x) = p(x+a) - p(x)$ which is a polynomial of lower degree in x. In short:

   $$p(x) = p(x+a) - \mathrm{D}_a p(x) \approx q(x+a) - \mathrm{D}_a p(x),$$

   where $\approx$ denotes nontrivial correlation. Over $\mathbb{F}_2$, this means that $p(x)$ equals $q(x+a) - \mathrm{D}_a p(x)$ with probability $1/2 + \epsilon$ over a. Thus we can compute $p(x)$ with high probability by taking the majority over several random choices of a. This analysis does not quite work over larger fields, and seems to require additional ideas.

On the seed length of our generator. For fixed ϵ, d, and $\mathbb{F}$, known constructions of generators for linear polynomials give seed length $d \cdot \log_{|\mathbb{F}|} n + O(1)$ for our generator. It is natural to ask whether the seed length can be improved, say by using fewer than d independent copies of the linear generator. In the full version we show this is not possible: To fool degree-d polynomials, seed length $d \cdot \log n$ is required. In particular, generators for linear polynomials in general do not fool quadratic polynomials.

Proposition 3. *For every $\mathbb{F}$, d and ϵ and sufficiently large n, there exists a linear generator that ϵ-fools linear polynomials over $\mathbb{F}^n$ but such that the sum of $d-1$ independent*

copies of the generator does not (0.9)*-fool degree* d *polynomials over* $\mathbb{F}^n$.

The results by Shachar Lovett. Subsequently to our results, Shachar Lovett [13] showed unconditionally that the sum of $2^{O(d)}$ generators that $\epsilon^{2^{O(d)}}$-fool linear functions fools degree-d polynomials with error ϵ. Some of the ideas in [13] can also be used in our analysis to obtain better parameters for $d \geq 3$.

Organization. This paper is organized as follows. In Section 3 we deal with the case of small $U_d(p)$, in Section 4 with the case of large $U_2(p)$, and in Section 5 we show our generator for quadratic polynomials. In Section 6 we deal with the case of large $U_d(p)$ and show how to self-correct polynomials. In Section 7 we show our generators for higher degree polynomials.

2 Preliminaries

In this paper we work over arbitrary prime fields $\mathbb{F}$. Although all our main ideas are already present in the results for the fundamental case $\mathbb{F}_2 = \{0,1\}$, we will obtain results about arbitrary prime fields at essentially no additional cost, and thus we present our results in generality. For this generality, it is useful to introduce the following notation.

Notation 4. *For a prime field* $\mathbb{F}$ *and* $x \in \mathbb{F}$*, we denote by* $e(x) \in \mathbb{C}$ *the value* ω^x*, where* ω *is the primitive root of unity* $e^{2\pi \mathbf{i}/|\mathbb{F}|}$*. The field will always be clear from the context. For a random variable* $X \in \mathbb{F}$*, we extensively use the notation*

$$\mathrm{E}_X\, e\,[X] := E_X[e(X)].$$

The notion of pseudorandomness that we use is the following.

Definition 5 (Pseudorandomness)**.** *We say that a distribution* W *on* $\mathbb{F}^n$ fools *degree-*d *polynomials with error* ϵ *if for every degree-*d *polynomial* p *we have*

$$|\mathrm{E}_{X\in\mathbb{F}^n}\, e\,[p(X)] - \mathrm{E}_W\, e\,[p(W)]| \leq \epsilon.$$

A generator $G : \mathbb{F}^s \to \mathbb{F}^n$ fools *degree-*d *polynomials with error* ϵ *if the distribution* $G(X)$ *does (for random* $X \in \mathbb{F}^s$*).*

Remark 6 (On Definition 5)**.** *We point out that pseudorandomness is often defined in terms of* statistical distance*. However, the algebraic Definition 5 is more convenient for the purposes in this paper. Our results are easily seen to be equivalent to results in terms of statistical distance, and this is formally proved in the full version of this paper.*

The basic building block of our construction is a generator for degree-1 polynomials. This generator was first obtained for $\mathbb{F}_2$ in [15]. Then [2] gave other constructions over $\mathbb{F}_2$, and it has since been observed by several researchers that constructions exist over any prime field. In particular, we have the following.

Lemma 7 ([15], Proposition 4.1 in [6])**.** *For every* ϵ*, prime field* $\mathbb{F}$*, and sufficiently large* n*, there is an explicit generator* $G : \mathbb{F}^s \to \mathbb{F}^n$ *that fools linear polynomials with error* ϵ *with seed length* $s = c \cdot \log_{|\mathbb{F}|}(n/\epsilon)$*, where* c *is an absolute constant.*

By *explicit* in the above lemma we mean that given an input seed and an index $i \leq n$, the i-th output field element can be computed in time polynomial in $|\mathbb{F}|$ and s. We note that the construction in Proposition 4.1 in [6] is a straightforward extension of the "powering" construction in [2] to larger fields. This construction requires to find an irreducible polynomial of degree s over $\mathbb{F}$, which can be done in time polynomial in $\mathbb{F}$ and s [18]. If such a polynomial is given or preprocessed, then the generator is computable in time polynomial in $\log_2 |\mathbb{F}|$ and s.

2.1 The degree norm

In this section we discuss the degree-d norm. For a function $f : \mathbb{F}^n \to \mathbb{F}$, we define its *directional derivative* $\mathrm{D}_y f$ in the direction of $y \in \mathbb{F}^n$ to be the function

$$\mathrm{D}_y f(x) := f(x+y) - f(x).$$

When f is a polynomial of degree d, all its directional derivatives are polynomials of degree at most $d-1$. We can take multiple derivatives too: We write $\mathrm{D}_{y_1,\ldots,y_k} f(x)$ for the function

$$\mathrm{D}_{y_1} \ldots \mathrm{D}_{y_k} f(x) = \sum\nolimits_{S\subseteq[k]} (-1)^{k-|S|} f\left(x + \sum\nolimits_{i\in S} y_i\right),$$

which we call a derivative of order k. If f is a polynomial of degree d, this will be a polynomial of degree at most $d-k$, and this does not depend on the order in which the derivatives are taken. The following claim, which is not hard to verify, states this formally.

Fact 8. *For every polynomial* $f : \mathbb{F}^n \to \mathbb{F}$ *of degree* d *and every* $y_1, \ldots, y_k \in \mathbb{F}^n$*, the function* $x \to \mathrm{D}_{y_1,\ldots,y_k} f(x)$ *is a polynomial of degree* $d-k$*.*

We now give the definition of the norm. Although this is syntactically defined as the expectation of a complex-valued random variable, it is always a non-negative real number (see, e.g., [17]).

Definition 9 (Degree-k norm[4]). *Let $f : \mathbb{F}^n \to \mathbb{F}$ be a function and $k \geq 1$ an integer. The* degree-k norm *of f is defined as*

$$U_k(f) := \mathrm{E}_{Y_1,Y_2,\ldots,Y_k,X \in \mathbb{F}^n} e\left[\mathrm{D}_{Y_1,\ldots,Y_k} f(X)\right].$$

3 When $U_d(f)$ is small: Fooling the Gowers norm

In this section we prove a generalization of the following fact relating the bias of a polynomial to its degree norm (cf. [21, Lemma 2.3]): For every degree-d polynomial $p : \mathbb{F}^n \to \mathbb{F}$ over a prime field $\mathbb{F}$,

$$\left|\mathrm{E}_X e\left[p(X)\right]\right| \leq U_d(p)^{1/2^d}. \tag{4}$$

The generalization that we prove is the following.

Lemma 10. *Let $p : \mathbb{F}^n \to \mathbb{F}$ be a degree-d polynomial over a prime field $\mathbb{F}$. Let $W_1, \ldots, W_d$ be d independent distributions that fool linear tests over $\mathbb{F}$ with error ϵ. Then*

$$\left|\mathrm{E}_{W_1,\ldots,W_d} e\left[p(W_1 + \cdots + W_d)\right]\right| \leq \left(U_d(p) + d \cdot \epsilon\right)^{1/2^d}.$$

Before discussing the proof of Lemma 10 we make some remarks.

Remark 11. *We observe the following. (1) Lemma 10 indeed generalizes Fact (4), because the uniform distribution fools linear tests with error $\epsilon = 0$, and XOR'ing together independent uniform distributions simply results in the uniform distribution. (2) Lemma 10 shows that the distribution $W_1 + \cdots + W_d$ fools degree-d polynomials* if *their degree-d norm is small. This is because in this case both $\left|\mathrm{E}_X e\left[p(X)\right]\right|$ and $\left|\mathrm{E}_{W_1,\ldots,W_d} e\left[p(W_1 + \cdots + W_d)\right]\right|$ are small, and so is their difference by the triangle inequality.*

We now discuss the proof of Lemma 10. For the proof we need some claims.

Definition 12. *A function $f : \mathbb{F}^{n \times d} \to \mathbb{F}$ in variables $y_{1,1}, \ldots, y_{d,n}$ is* block-linear *if for every i, it is a linear function of the variables $y_{i,1}, \ldots, y_{i,n}$.*

Claim 13. *Let $p : \mathbb{F}^n \to \mathbb{F}$ be a degree-d polynomial over a prime field $\mathbb{F}$. Then the function $q_p(y_1, \ldots, y_d) := \mathrm{D}_{y_1,\ldots,y_d} p(x)$ is block-linear.*[5].

[4] In general the degree-k norm is defined for functions $\mathbb{F}^n \to \mathbb{C}$. Functions from $\mathbb{F}^n$ to $\mathbb{C}$ form a vector space over $\mathbb{C}$ and the degree-d norm is indeed a norm of this space when raised to the power of $1/2^d$; see, e.g., [11].

[5] Note that $\mathrm{D}_{y_1,\ldots,y_d} f(x)$ does not depend on x anymore because we are taking d derivatives of a degree-d polynomial; see Section 2.1

Claim 14. *For every function $f : \mathbb{F}^n \to \mathbb{F}$ and every distribution $\mathcal{D}$ over $\mathbb{F}^n$,*

$$\left|\mathrm{E}_{W_1,\ldots,W_d} e\left[f(W_1 + \cdots + W_d)\right]\right|^{2^d} \leq \mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}} e\left[\mathrm{D}_{W'_1 - W_1,\ldots,W'_d - W_d} f(W_1 + \cdots + W_d)\right],$$

where $W_1, \ldots, W_d, W'_1, \ldots, W'_d$ are independent samples from $\mathcal{D}$.

Before proving the above claims, let us see why they imply the main result of this section, Lemma 10.

Proof of Lemma 10 from Claims 13 and 14. Let $q_p(y_1, \ldots, y_d) := \mathrm{D}_{y_1,\ldots,y_d} p(x)$ and $W := W_1 + \cdots + W_d$. We have:

$$\begin{aligned}
&\mathrm{E}_{W_1,\ldots,W_d} e\left[p(W)\right]^{2^d} \\
&\leq \mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}} e\left[\mathrm{D}_{W'_1 - W_1,\ldots,W'_d - W_d} p(W)\right] \\
&= \mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}} e\left[q_p(W'_1 - W_1, \ldots, W'_d - W_d)\right] \\
&\leq \mathrm{E}_{Y_1,\ldots,Y_d \in \mathbb{F}^n} e\left[q_p(Y_1, \ldots, Y_d)\right] + \epsilon \cdot d \qquad (5) \\
&= U_d(p) + \epsilon \cdot d,
\end{aligned}$$

which proves the lemma, except for Inequality (5) which we now justify. The inequality holds because q_p is a block-linear polynomial by Claim 13, and each of the $W'_i - W_i$ fools linear tests with error ϵ. Specifically, letting H_i denote the i-th hybrid

$$H_i := Y_1, \ldots, Y_i, W'_{i+1} - W_{i+1}, \ldots, W'_d - W_d$$

we have

$$\begin{aligned}
&\mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}} e\left[q_p(W'_1 - W_1, \ldots, W'_d - W_d)\right] \\
&\qquad - \mathrm{E}_{Y_1,\ldots,Y_d} e\left[q_p(Y_1, \ldots, Y_d)\right] \\
&= \mathrm{E}_{H_0} e\left[q_p(H_0)\right] - \mathrm{E}_{H_d} e\left[q_p(H_d)\right] \\
&= \sum_{i=0}^{d-1} \mathrm{E}_{H_i} e\left[q_p(H_i)\right] - \mathrm{E}_{H_{i+1}} e\left[q_p(H_{i+1})\right] \\
&\leq \epsilon \cdot d. \qquad \square
\end{aligned}$$

Proof of Claim 13. The claim follows from the fact that taking j derivatives decreases the degree by j (Fact 8), and the fact that the operation of taking derivatives is invariant under different orders for the directions, i.e. taking a derivative with respect to y and then another one with respect to z is the same as taking a derivative with respect to z and then another one with respect to y. This latter fact follows easily from the definition of derivative in Section 2.1. (See Fact 17 for examples.) $\square$

Proof of Claim 14. We proceed by induction on d. When $d = 1$ we have

$$\begin{aligned}\left|\mathrm{E}_{W_1} e\left[f(W_1)\right]\right|^2 &= \mathrm{E}_{W_1,W_1'} e\left[f(W_1') - f(W_1)\right] \\ &= \mathrm{E}_{W_1,W_1'} e\left[\mathrm{D}_{W_1'-W_1} f(W_1)\right].\end{aligned}$$

For $d > 1$, let $W := W_1 + \cdots + W_d$. Using the fact that $|\mathrm{E}_Z[Z]|^2 \le \mathrm{E}_Z[|Z|^2]$ for any complex random variable Z, we have

$$\begin{aligned}&\left|\mathrm{E}_{W_1,\ldots,W_d} e\left[f(W)\right]\right|^{2^d} \\ &\le \mathrm{E}_{W_1,\ldots,W_{d-1}}\left[\left|\mathrm{E}_{W_d} e\left[f(W)\right]^2\right|^{2^{d-1}}\right] \\ &= \mathrm{E}_{W_1,\ldots,W_{d-1}}\left[\mathrm{E}_{W_d,W_d'} e\left[\mathrm{D}_{W_d'-W_d} f(W)\right]\right]^{2^{d-1}} \\ &\le \mathrm{E}_{W_d,W_d'}\left[\left|\mathrm{E}_{W_1,\ldots,W_{d-1}} e\left[\mathrm{D}_{W_d'-W_d} f(W)\right]\right|^{2^{d-1}}\right] \\ &\le \mathrm{E}_{\substack{W_d\\W_d'}}\left[\mathrm{E}_{\substack{W_1,\ldots,W_{d-1}\\W_1',\ldots,W_{d-1}'}} e\left[\mathrm{D}_{W_1'-W_1,\ldots,W_d'-W_d} f(W)\right]\right] \\ &= \mathrm{E}_{\substack{W_1,\ldots,W_d\\W_1',\ldots,W_d'}} e\left[\mathrm{D}_{W_1'-W_1,\ldots,W_d'-W_d} f(W)\right].\end{aligned}$$

The third line follows from the case $d = 1$, while the fifth line follows from the inductive hypothesis. □

4 When $U_2(p)$ is large: Canonical forms of quadratics

In this section we prove the following lemma.

Lemma 15. *Let $p : \mathbb{F}^n \to \mathbb{F}$ be a quadratic polynomial over a prime field $\mathbb{F}$. Let W be a distribution that fools linear polynomials with error ϵ. Then*

$$\left|\mathrm{E}_X e\left[p(X)\right] - \mathrm{E}_W e\left[p(W)\right]\right| = O\big(\epsilon / U_2(p)\big).$$

Note that Lemma 15 shows that the distribution W fools degree-2 polynomials p *if* their degree-2 norm is large. The proof of the lemma is based on two other results, discussed in the next subsections.

4.1 Canonical representations of quadratic polynomials

In this subsection we prove the following lemma.

Lemma 16. *Every quadratic polynomial p over a prime field $\mathbb{F}$ is a function of $\log_{|\mathbb{F}|}(1/U_2(p)) + 1$ linear functions.*

To get a sense of the parameters, note that if p is linear then $U_2(p) = 1$ and indeed p is a function of 1 linear function, namely p itself.

To prove the lemma we make use of a some auxiliary results. The following Fact lists some simple and useful properties of the degree norm. The proofs follow easily from the definition of the degree norm, and are given at the end of this section.

Fact 17 (Properties of the degree norm)**.** *Let $\mathbb{F}$ be a prime field and let $f : \mathbb{F}^n \to \mathbb{F}$ be a function. The following hold.*

1. *For a function $f' : \mathbb{F}^{n'} \to \mathbb{F}$, let $(f + f') : \mathbb{F}^{n+n'} \to \mathbb{F}$ be the function $(f + f')(x, x') := f(x) + f'(x')$. Then $U_k(f + f') = U_k(f) \cdot U_k(f')$.*
2. *For an invertible matrix A, let $f \circ A : \mathbb{F}^n \to \mathbb{F}$ denote the function $(f \circ A) := f(Ax)$. Then $U_k(f) = U_k(f \circ A)$.*
3. *For every polynomial q of degree $k - 1$ or less, $U_k(f + q) = U_k(f)$.*
4. *Let $f : \mathbb{F} \to \mathbb{F}$, $|\mathbb{F}|$ odd, be $f(x) := a \cdot x^2$ for $a \in \mathbb{F}, a \neq 0$. Then $U_2(f) = \mathrm{E}_{Y_1,Y_2 \in \mathbb{F}} e\left[2 \cdot a \cdot Y_1 \cdot Y_2\right] = 1/|\mathbb{F}|$.*
5. *Let $f : \mathbb{F}_2^2 \to \mathbb{F}_2$ be $f(x_1, x_2) := x_1 \cdot x_2$. Then $U_2(f) = \mathrm{E}\, e\left[Y_{11} \cdot Y_{22} + Y_{21} \cdot Y_{12}\right] = 1/4$.*

The next lemma is a standard result in the theory of quadratic forms according to which every quadratic form is equivalent, up to an invertible linear transformation of the variables, to a quadratic form where no two quadratic monomials share a variable. The statement (and proof) of these results differ according to whether the field has even size or not.

Lemma 18 (Theorems 6.21 and 6.30 in [12])**.** *For every quadratic polynomial $p : \mathbb{F}^n \to \mathbb{F}$ over a prime field $\mathbb{F}$ there exists an invertible matrix A, a linear polynomial ℓ, and field elements $c_1, c_2, \ldots, c_n$ (some of which may be 0) such that:*

1. *If $q = 2$ then $(p \circ A)(x) = \sum_{1 \le i \le \lfloor n/2 \rfloor} c_i \cdot x_{2i-1} \cdot x_{2i} + \ell(x)$,*
2. *If q is odd then $(p \circ A)(x) = \sum_{1 \le i \le n} c_i \cdot x_i^2 + \ell(x)$.*

Armed with the above results, we now present the proof of Lemma 16.

Proof of Lemma 16. We present the proof for the case in which $|\mathbb{F}|$ is an odd prime because this case shows the dependence on the field size and the case $\mathbb{F} = \mathbb{F}_2$ is analogous. By Lemma 18, there exists an invertible matrix A, a linear polynomial ℓ, and field elements $c_1, c_2, \ldots, c_n$ (some of which may be 0) such that: $(p \cdot A)(x) = \sum_{1 \le i \le n} c_i \cdot x_i^2 + \ell(x)$. Let us assume without loss of generality that exactly the first s $c_i's$ are non-zero, i.e. $c_i = 0$ if and only if $i > s$.

We now have:

$$\begin{aligned}
\epsilon &\leq U_2(p) && \text{by assumption}\\
&= U_2\left(\sum\nolimits_{1\leq i\leq n} c_i\cdot x_i^2 + \ell(x)\right) && \text{by Fact 17.2}\\
&= U_2\left(\sum\nolimits_{1\leq i\leq s} c_i\cdot x_i^2\right) && \text{by Fact 17.3}\\
&= \prod\nolimits_{1\leq i\leq s} U_2\left(c_i\cdot x_i^2\right) && \text{by Fact 17.1}\\
&= 1/|\mathbb{F}|^s && \text{by Fact 17.4.}
\end{aligned}$$

The above derivation shows that, up to a linear transformation A, the original polynomial p is equivalent to a quadratic polynomial where at most $s = \log_{|\mathbb{F}|}(1/\epsilon)$ variables appear in a degree-2 monomial. Taking into account the linear part ℓ, we have that p is a function of at most $t := s + 1$ linear polynomials which proves the lemma. □

Proof of Fact 17. Item (1) follows by linearity of the directional derivative operator and statistical independence of the variables of f and f'. For item (2) we have, taking expectations over $Y_1, Y_2, \ldots, Y_k, X \in \mathbb{F}^n$,

$$\begin{aligned}
&U_k(f\circ A)\\
&= \mathrm{E}\, e\left[\sum\nolimits_{S\subseteq[k]}(-1)^{k-|S|} f\circ A(X+\sum\nolimits_{i\in S} Y_i)\right]\\
&= \mathrm{E}\, e\left[\sum\nolimits_{S\subseteq[k]}(-1)^{k-|S|} f(AX+\sum\nolimits_{i\in S} AY_i)\right]\\
&= \mathrm{E}\, e\left[\sum\nolimits_{S\subseteq[k]}(-1)^{k-|S|} f(X+\sum\nolimits_{i\in S} Y_i)\right]\\
&= U_k(f).
\end{aligned}$$

Item (3) follows from the fact that q vanishes after taking k directional derivatives. The second equality in Item (4) is justified as follows. Whenever $Y_1 \neq 0$, the value $2\cdot a\cdot y_1\cdot y_2$ is uniform, for random Y_2, over the $|\mathbb{F}|$ complex roots of unity (recall that $a \neq 0$). Consequently, in this case the expectation is 0. Since $Y_1 \neq 0$ with probability $1 - 1/|\mathbb{F}|$, and when $Y_1 = 0$ the expectation, for random Y_2, is 1, the last inequality above is indeed justified. Item (5) is proved analogously. □

4.2 Fooling functions of few linear polynomials

In this subsection we show that a distribution that fools linear polynomials also fools functions of few such polynomials. This, in combination with Lemma 16 proves the main Lemma 15 of this section.

We need some basic Fourier analysis, which we now briefly recall. Every function $h : \mathbb{F}^k \to \mathbb{C}$ can be written in the form $h(x) = \sum_{a\in\mathbb{F}^n} \hat{h}_a\chi_a(x)$, where χ_a is the function $\chi_a(x) := e(a_1x_1 + \cdots + a_kx_k)$ and the *Fourier coefficients* $\hat{h}_a$ are given by $\hat{h}_a = E_x\big[h(x)\cdot\overline{\chi_a(x)}\big]$. The Fourier coefficients satisfy Parseval's identity

$$\sum\nolimits_{a\in\mathbb{F}^k}|\hat{h}_a|^2 = \frac{1}{|\mathbb{F}|^k}\sum\nolimits_{x\in\mathbb{F}^k}|h(x)|^2.$$

We state the following lemma in more generality because it will be used for higher degrees later on. For now we are only interested in the case when $\mathcal{F}$ is the class of linear polynomials.

Lemma 19. *Suppose that W is ϵ-pseudorandom for any class $\mathcal{F}$ of functions $\mathbb{F}^n \to \mathbb{F}$ that forms a linear space. Then for every function $h : \mathbb{F}^k \to \mathbb{C}$ and every collection of functions $f_1, \ldots, f_k \in \mathcal{F}$,*

$$\begin{aligned}
\big|&\mathrm{E}_{X\sim\mathbb{F}^n}[h(f_1(X),\ldots,f_k(X))]\\
&\qquad - \mathrm{E}_W[h(f_1(W),\ldots,f_k(W))]\big| \leq \epsilon\cdot L(h),
\end{aligned}$$

where $L(h) := \sum_{a\in\mathbb{F}^n}|\hat{h}_a|$ is the ℓ_1 norm of (the Fourier transform of) h.

Before proving Lemma 19, let us see how it can be used to prove the main Lemma 15 of this section. We are going to use the following trivial bound on the ℓ_1 norm of a function, which just uses the number of variables it depends on.

Proposition 20. *Let $h : \mathbb{F}^k \to \mathbb{C}$ be a function such that $|h(x)| = 1$. Then $L(h) \leq |\mathbb{F}|^{k/2}$.*

Proof. By Cauchy-Schwarz and Parseval's identity, we have that

$$L(h) = |\mathbb{F}|^k\,\mathrm{E}_a[\hat{h}_a] \leq |\mathbb{F}|^k\sqrt{\mathrm{E}_a[\hat{h}_a^2]} = |\mathbb{F}|^{k/2}. \qquad \square$$

Proof of Lemma 15 assuming Lemma 19. By Lemma 16 we have that $p = f(\ell_1(x),\ldots,\ell_t(x))$ where the ℓ's are linear polynomials, $f : \mathbb{F}^t \to \mathbb{F}$ and $t = O(\log_{|\mathbb{F}|}(1/U_2(p)))$. By Lemma 19 and Proposition 20 we have that

$$|\mathrm{E}_X\, e[p(X)] - \mathrm{E}_W\, e[p(W)]| \leq |\mathbb{F}|^t\cdot\epsilon = O\left(\frac{\epsilon}{U_2(p)}\right).$$

□

It now only remains to prove Lemma 19.

Proof of Lemma 19. For every distribution Y on $\mathbb{F}^k$, we have that

$$\mathrm{E}_Y[h(Y)] = \sum\nolimits_{a\in\mathbb{F}^k}\hat{h}_a\,\mathrm{E}_Y[\chi_a(Y)]. \tag{6}$$

Now fix $a \in \mathbb{F}^k$, and consider the case $Y = (f_1(Z),\ldots,f_k(Z))$, where Z is now sampled from some distribution on $\mathbb{F}^n$. We have that

$$\begin{aligned}
\mathrm{E}_Y[\chi_a(Y)] &= \mathrm{E}_Y\, e[a_1Y_1 + \cdots + a_kY_k]\\
&= \mathrm{E}_Z\, e[a_1f_1(Z) + \cdots + a_kf_k(Z)].
\end{aligned}$$

By linearity, $a_1 f_1 + \cdots + a_k f_k \in \mathcal{F}$, so because W is pseudorandom for $\mathcal{F}$ we have

$$\left|\mathrm{E}_X[\chi_a(X)] - \mathrm{E}_W[\chi_a(W)]\right| \leq \epsilon.$$

where X is uniformly random in $\mathbb{F}^k$ and W is pseudorandom. By equation (6),

$$\begin{aligned}&\left|\mathrm{E}_X\, e\left[h(X)\right] - \mathrm{E}_W\, e\left[h(W)\right]\right| \\ &\quad\leq \sum\nolimits_{a\in\mathbb{F}^k} \hat{h}_a \left|\mathrm{E}_X\, e\left[h(X)\right] - \mathrm{E}_W\, e\left[h(W)\right]\right| \\ &\quad\leq \epsilon \cdot \sum\nolimits_{a\in\mathbb{F}^k} \hat{h}_a = \epsilon \cdot L(h).\end{aligned}$$

□

5 Fooling quadratic polynomials

Theorem 21 (Pseudorandom generators for quadratic polynomials)**.** *There is an absolute constant c such that the following holds for every n and prime field $\mathbb{F}$. Let W_1, W_2 be 2 independent distributions over $\mathbb{F}^n$ that fool linear tests with error $1/n^c$. Then $W := W_1 + W_2$ fools quadratic tests with error $1/n$: For every quadratic polynomial p over $\mathbb{F}$ we have*

$$\left|\mathrm{E}_{X\in\mathbb{F}^n}\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \leq 1/n.$$

In particular, there exists an efficiently computable generator $G : \mathbb{F}^s \to \mathbb{F}^n$ that fools to within $1/n$ quadratic polynomials over $\mathbb{F}$, where $\mathbb{F}$ is a prime field, with seed length $s = O(\log_{|\mathbb{F}|} n)$.

Proof. We argue by case analysis, according to the value of $U_2\left(p\right)$. Let $\tau := 1/n$ (note that the error of the W's is τ^c).

Case $U_2\left(p\right) \leq \tau^{c/2}$**.** In this case we have

$$\begin{aligned}&\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \\ &\quad\leq \left|\mathrm{E}_X\, e\left[p(X)\right]\right| + \left|\mathrm{E}_W\, e\left[p(W)\right]\right| \\ &\quad\leq (\tau^{c/2})^{1/4} + (\tau^{c/2} + 2\cdot\tau^c)^{1/4}.\end{aligned}$$

The first line is the triangle inequality, while the second line uses Lemma 10 twice. For sufficiently large c this concludes the case.

Case $U_2\left(p\right) \geq \tau^{c/2}$**.** In this case Lemma 15 gives

$$\begin{aligned}\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| &= O(\tau^c / U_2\left(p\right)) \\ &= O(\tau^{c/2}).\end{aligned}$$

The above case analysis concludes the proof of the theorem, except for the "in particular" part. This part follows by taking the sum of two independent linear generators (Lemma 7). □

6 When $U_d\left(p\right)$ is large: Correlation amplification

In this section we prove a lemma which is similar to Lemma 15 but works for higher degrees. For degree 3 we obtain an unconditional result, while for higher degrees the lemma relies on a special case of the "Gowers inverse conjecture," which we now describe.

Conjecture 22 (d vs. $d-1$ Gowers inverse conjecture over $\mathbb{F}$)**.** *For every $\tau \geq 0$ there exists a real number $IG^d_{\mathbb{F}}(\tau) > 0$ such that for every n and every polynomial $p : \mathbb{F}^n \to \mathbb{F}$ of degree d the following is true:*

$$\begin{aligned}&U_d\left(p\right) \geq \tau \\ &\Rightarrow \max_{\substack{q:\,\mathbb{F}^n\to\mathbb{F} \\ \deg(q)=d-1}} \left|\mathrm{E}_X\, e\left[p(X) - q(X)\right]\right| \geq IG^d_{\mathbb{F}}(\tau).\end{aligned}$$

Remark 23. *Conjecture 22 is usually stated for* arbitrary *functions p [11, 16]. However, for the results in this work the special case where p has degree d is sufficient. We call this the d vs. $d-1$ Gowers inverse conjecture. It remains to be seen whether this case (d vs. $d-1$) is more difficult than the general one (n vs. $d-1$).*

The following lemma says that, if we believe the d vs. $d-1$ Gowers inverse Conjecture 22, distributions that fool polynomials of degree $d-1$ also fool polynomials of degree d provided their degree-d norm is large. We write $IG(\tau)$ for $IG^d_{\mathbb{F}}(\tau)$ when $\mathbb{F}$ and d are clear from the context.

Lemma 24. *Let $p : \mathbb{F}^n \to \mathbb{F}$ be a degree-d polynomial over a prime field $\mathbb{F}$. Let W be a distribution that fools degree $d-1$ polynomials with error ϵ. Assume that $U_d\left(p\right) \geq \tau$ and that Conjecture 22 holds for $\mathbb{F}$ and d. Then*

$$\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \leq \exp(B)\cdot\epsilon^{1/B},$$

where $B \leq (|\mathbb{F}|/IG(\tau))^c$ for a universal constant $c > 0$. $p : \mathbb{F}^n \to \mathbb{F}$ be a degree-d polynomial over a prime field $\mathbb{F}$. Let W be a distribution that fools degree $d-1$ polynomials with error ϵ. Assume that $U_d\left(p\right) \geq \tau$ and that Conjecture 22 holds for $\mathbb{F}$ and d. Then

To prove Lemma 24, we would like a stronger statement than what is given to us by the Gowers inverse conjecture. The conjecture says that if $U_d(p) > \alpha$ then p has correlation at least α' with a degree $d-1$ polynomial. We would like the stronger property that p has correlation $(1-\epsilon)$ with a degree $d-1$ polynomial. Although this is not true, we will show that p does have correlation $(1-\epsilon)$ with a function of few degree $d-1$ polynomials. This will be sufficient to prove Lemma 24 using Lemma 19.

6.1 Amplifying correlation via self-correction

The following lemma shows that if a function f is somewhat correlated to another function g, then f is highly correlated with some function of evaluations of g minus a derivative of f. In particular, if f has degree d and g has degree $d-1$ then f has high correlation with a function that depends on a small number of degree $d-1$ polynomials.

Lemma 25. *Let $f, g : \mathbb{F}^n \to \mathbb{F}$, for a prime field $\mathbb{F}$. Suppose that $|\mathrm{E}_{x\in\mathbb{F}^n} e[f(x)-g(x)]| \geq \delta$. Then there is a in integer t and a function $B\colon \mathbb{F}^t \to \mathbb{F}$ such that for every $x \in \mathbb{F}^n$*

$$\mathrm{Pr}_{a_1,a_2,\ldots,a_t\in\mathbb{F}^n}\big[f(x) \neq B\big(g(x+a_1)-\mathrm{D}_{a_1} f(x), \\ ,\ldots,g(x+a_t)-\mathrm{D}_{a_t} f(x)\big)\big] < \gamma,$$

where $t \leq (|\mathbb{F}|/\delta)^{O(1)}(1+\log 1/\gamma)$.

Lemma 25 is relatively easy to obtain over $\mathbb{F}_2$, and in this case B can be taken to be the majority function (or its negation). A somewhat more involved argument seems necessary for larger fields.

The complete proof of Lemma 25 is given in the full version of this paper; here we just give the following intuition.

Intuition for the proof of Lemma 25. Let us fix x and look at the following identity, which holds for every $a \in \mathbb{F}^n$:

$$f(x) = f(x+a) - \mathrm{D}_a f(x).$$

If we think of the right hand side as a function of a, this identity tells us that this function always evaluates to the constant $f(x) = r$. Now suppose that instead of having access to the function on the right, we can only look at the "corrupted" version

$$v(a) := g(x+a) - \mathrm{D}_a f(x) = r + (g(x+a) - f(x+a)). \tag{7}$$

Let us look at this as a coding problem: Elements of $\mathbb{F}$ are encoded by functions from $\mathbb{F}^n$ to $\mathbb{F}$. The valid encoding of every $r \in \mathbb{F}$ is the constant function r. Instead of r, the decoder has access to some corrupted $v : \mathbb{F}^n \to \mathbb{F}$. We know that v satisfies Equation (7) and, by our assumption, that $|\mathrm{E}_{x\in\mathbb{F}^n} e[f(x)-g(x)]| \geq \delta$. Can the decoder recover r using a small number of queries to the corrupted codeword? If so, then the decoder is a function that approximates $f(x)$ in terms of a few copies of $g(x+a) - \mathrm{D}_a f(x)$. We show that this recovery is indeed possible provided that the decoder is given some side information, specifically the distribution of values of $g - f : \mathbb{F}^n \to \mathbb{F}$.

First, we need to see what the corrupted codeword will look like. Let us view $g-f$ as a function from $\mathbb{F}^n \to \mathbb{F}$. Let p denote the distribution of values of $g-f$ over $\mathbb{F}$ (namely $p(s) := \mathrm{Pr}_a[g(a)-f(a)=s]$).

$$\begin{aligned}\mathrm{Pr}_a[v(a)=s] &= \mathrm{Pr}_a[g(x+a)-f(x+a)=s-r] \\ &= \mathrm{Pr}_a[g(a)-f(a)=s-r] \\ &= p(s-r) = p_{-r}(s),\end{aligned} \tag{8}$$

where p_{-i} is the "shifted" distribution that assigns to $y-i$ the probability $p(y)$, namely $p_{-i}(y-i) = p(y)$.

The key observation is that the value r is uniquely determined by equation (8). For suppose that there were some $r' \neq r$ such that $p_{-r'}(s) = \mathrm{Pr}_a[v(a)=s]$ for all $s \in \mathbb{F}$. Then p_{-r} and $p_{-r'}$ are the same distribution, and it is not difficult to see that this is only possible if p is the uniform distribution, which can be shown to contradict our assumption that $|\mathrm{E}_{x\in\mathbb{F}^n} e[f(x)-g(x)]| \geq \delta$.

This suggests a natural approach to the decoding problem: On input x, compute the probabilities $\mathrm{Pr}_a[v(a)=s]$ for every $s \in \mathbb{F}$ and decode to the unique r that satisfies equation (8). This is infeasible, as to compute the probabilities exactly we have to query v everywhere, but we can obtain very good estimates by sampling. We will need to argue that if the estimates are sufficiently good, then r is still uniquely determined. Specifically we argue that r is the value that minimizes the *statistical distance* between the distribution $p_{-r}(s)$ and the empirical distribution of the sample.

6.2 Proof of Lemma 24

We now prove Lemma 24.

Proof of Lemma 24. By assumption and Conjecture 22, there is a polynomial q of degree $d-1$ such that

$$|\mathrm{E}_X e[p(x)-q(x)]| \geq \delta,$$

where $\delta := IG(\tau)$.

To avoid notational clutter, let us set $\eta := \delta/|\mathbb{F}|$. We now set $\gamma := \epsilon^{\eta^c}$, for a sufficiently large universal constant c to be determined later, and let

$$h(a,x) := B(q(x+a_1) - D_{a_1} f(x), \\ ,\ldots,q(x+a_t) - D_{a_t} f(x))$$

be the function in Lemma 25 where $a := (a_1, a_2, \ldots, a_t)$. First, we argue that $h(A,Y)$ approximates $p(Y)$ both when Y is random and when Y is pseudorandom. In fact, let Y be an arbitrary distribution. Then

$$\begin{aligned}&\big|\mathrm{E}_Y e[p(Y)] - \mathrm{E}_{A,Y} e[h(A,Y)]\big| \\ &\quad\leq \mathrm{E}_Y \mathrm{E}_A\big[|e(p(Y)) - e(h(A,Y))|\big] \\ &\quad\leq \mathrm{E}_Y\big[2\cdot \mathrm{Pr}_A[p(Y)\neq h(A,Y)]\big] \\ &\quad< 2\cdot\gamma,\end{aligned} \tag{9}$$

where the last inequality follows from Lemma 25.

Next, we argue that the bias of h over truly random X is close to the bias of h over pseudorandom W. For fixed a, $h(a,x)$ is a function of t degree $d-1$ polynomials $g(x+a_i)-\mathrm{D}_{a_i} f(x)$. Since degree $d-1$ polynomials form a linear space and W is ϵ-pseudorandom for such polynomials, by Lemma 19 and Proposition 20 we have that for every a,

$$\left|\mathrm{E}_X\, e\left[h(a,X)\right] - \mathrm{E}_W\, e\left[h(a,W)\right]\right| \le \epsilon \cdot |\mathbb{F}|^{t/2}. \quad (10)$$

and therefore

$$\begin{aligned}
&\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \\
&\quad\le \left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_{A,X}\, e\left[h(A,X)\right]\right| \\
&\qquad + \left|\mathrm{E}_W\, e\left[p(W)\right] - \mathrm{E}_{A,W}\, e\left[h(A,W)\right]\right| \\
&\qquad + \left|\mathrm{E}_{A,X}\, e\left[h(A,X)\right] - \mathrm{E}_{A,W}\, e\left[h(A,W)\right]\right| \\
&\quad\le 4\cdot\gamma + \mathrm{E}_A\left|\mathrm{E}_X\, e\left[h(A,X)\right] - \mathrm{E}_W\, e\left[h(A,W)\right]\right| \\
&\quad\le 4\cdot\gamma + \epsilon\cdot|\mathbb{F}|^{t/2},
\end{aligned}$$

where the second inequality follows from (9) and the third one is inequality (10). To conclude the proof, we only need to argue that the above error $4\cdot\gamma+\epsilon\cdot|\mathbb{F}|^{t/2}$ is of the desired form. Indeed, recalling that Lemma 25 gives $t \le (1/\eta)^{O(1)}(1+\log 1/\gamma)$ and that our choice of γ was $\gamma = \epsilon^{\eta^c}$, we have

$$\begin{aligned}
4\cdot\gamma+\epsilon\cdot|\mathbb{F}|^{t/2} &\le 4\cdot\epsilon^{\eta^c} + \epsilon\cdot|\mathbb{F}|^{(1/\eta)^{O(1)}(1+\eta^c\cdot\log 1/\epsilon)} \\
&\le 4\cdot\epsilon^{\eta^c} + \epsilon\cdot|\mathbb{F}|^{(1/\eta)^{O(1)}}\cdot\epsilon^{\eta^{c-O(1)}} \\
&\le \exp(\eta^{-c'})\cdot\epsilon^{\eta^{c'}},
\end{aligned}$$

for a sufficiently large universal constant c and another universal constant c'. □

7 Fooling higher degree polynomials

In this section we show that, assuming the Gowers inverse conjecture, polynomials of degree d in n variables can be ϵ-fooled by an explicit generator whose seed length is $O(d\cdot\log_{|\mathbb{F}|} n) + f(\epsilon,d,\mathbb{F})$, where f is independent on n. For $d=3$, the known Gowers inverse theorems [11, 16] yield unconditional generators with explicit estimates of the function f. We discuss the general d case and then we discuss the case $d=3$. We state the theorems in terms of distributions; the translation to the language of generators is simple.

Theorem 26. *Let a field $\mathbb{F}$ and a degree d be given. Assume Conjecture 22 for $\mathbb{F}$ and every degree $d' \le d$. Then for every $\epsilon>0$ there exists an $\epsilon_1>0$ such that if $W_1,\ldots,W_d$ are independent and fool linear polynomials on n inputs over $\mathbb{F}$ with error ϵ_1, then $W_1+\cdots+W_d$ fools degree-d polynomials on n inputs over $\mathbb{F}$ with error ϵ.*

Proof. We argue by induction on d. The base case $d=1$ holds trivially for $\epsilon=\epsilon_1$. Now let us assume that for every $\epsilon_{d-1}>0$ there exists $\epsilon_1>0$ such that the sum of d independent random variables that ϵ_1-fool linear tests on n inputs ϵ_{d-1}-fools degree $d-1$ tests on n inputs.

Let $W := W_1+\cdots+W_d$ and p be an arbitrary degree-d polynomial. We proceed analogously to Theorem 21, i.e. by case analysis according the the value of $U_d(p)$. Let $\tau := (\epsilon_d/4)^{2^d}$.

Case $U_d(p) \le \tau$. From Lemma 10 we have that

$$\begin{aligned}
&\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \\
&\quad\le \left|\mathrm{E}_X\, e\left[p(X)\right]\right| + \left|\mathrm{E}_W\, e\left[p(W)\right]\right| \\
&\quad\le \tau^{1/2^d} + (\tau+d\cdot\epsilon_1)^{1/2^d} \\
&\quad\le \epsilon_d.
\end{aligned}$$

where the last inequality holds by our choice of τ and sufficiently small ϵ_1.

Case $U_d(p) \ge \tau$. From Lemma 24 we have

$$\left|\mathrm{E}_X\, e\left[p(X)\right] - \mathrm{E}_W\, e\left[p(W)\right]\right| \le \exp(B)\cdot\epsilon_{d-1}^{1/B},$$

where $B \le (|\mathbb{F}|/IG(\tau))^{O(1)}$. Note that here we use the simple fact that W (ϵ_{d-1})-fools degree $d-1$ polynomials. (This is because for every degree $d-1$ polynomial $q(x)$ and every fixed $W_d = w_d$, the distribution $W_1+\cdots+W_{d-1}$ by assumption (ϵ_{d-1})-fools the polynomial $q(x+w_d)$.) To conclude the proof in this case, we only need to argue that $\exp(B)\cdot\epsilon_{d-1}^{1/B} \le \epsilon_d$ for sufficiently small ϵ_1. This is true because by inductive assumption ϵ_{d-1} goes to 0 when ϵ_1 does.

□

7.1 Fooling cubic polynomials

For $d=3$, the Gowers inverse conjecture was proved by Green and Tao [11] and Samorodnitsky [16], though with exponential slackness in the parameters.

Theorem 27 ([11, 16])**.** *Fix a prime field $\mathbb{F}$.*[6] *For $d=3$ and an absolute constant C, Conjecture 22 is true with*

$$IG^3_{\mathbb{F}}(\tau) \ge C\cdot\exp\left(-1/\tau^C\right).$$

Combining our approach with Theorem 27 we obtain the following unconditional generator for cubic polynomials.

[6] Green and Tao state their result for $\mathbb{F}_5$ but observe their argument extends over all odd prime fields.

Theorem 28. *Fix a prime field $\mathbb{F}$. Let $W_1, W_2, W_3 \in \mathbb{F}^n$ be independent distributions that $\exp(-2^{(2/\epsilon)^c})$-fool linear polynomials over $\mathbb{F}$, for a sufficiently large universal constant c. Then the distribution $W_1 + W_2 + W_3$ ϵ-fools degree 3 polynomials over $\mathbb{F}$.*

Proof. We follow the proof of Theorem 26. Let us think of being given $\epsilon = \epsilon_3$. In the proof of Theorem 26 the total error is dominated by the error in the case of large norm. In this case, the error is $\exp(B) \cdot \epsilon_2^{1/B}$, where $B \leq (|\mathbb{F}|/IG(\tau))^{O(1)}$. Since we are fixing the field, in the expression for B we can replace $|\mathbb{F}|$ with 2 (at the price of a different constant in the exponent). Recall that τ is polynomially related to ϵ_3. Using Theorem 27 we obtain that

$$B \leq C \cdot \exp\left(1/\epsilon_3^C\right)$$

for a universal constant C. Consequently, the error is at most ϵ_3 if

$$\exp(C \cdot \exp\left(1/\epsilon_3^C\right)) \cdot \epsilon_2^{1/\left(C \cdot \exp\left(1/\epsilon_3^C\right)\right)} \leq \epsilon_3.$$

The above inequality is satisfied for $\epsilon_2 = \exp(-\exp(1/\epsilon_3^{C'}))$ for a sufficiently large universal constant C', where ϵ_2 is such that $W_1 + W_2 + W_3$ ϵ_2-fools quadratic polynomials. The required bound on ϵ_2 follows by assumption (for sufficiently large c) using Theorem 21. □

Acknowledgments

We thank Anup Rao and Vladimir Trifonov for useful discussions about this problem in the initial stages of this work. We also thank Ben Green and Alex Samorodnitsky for helpful email exchanges about [11] and [16].

References

[1] N. Alon, J. Bruck, J. Naor, M. Naor, and R. M. Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on Information Theory*, 38(2):509–516, 1992.

[2] N. Alon, O. Goldreich, J. Håstad, and R. Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992.

[3] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn, and D. Ron. Testing low-degree polynomials over $\mathrm{GF}(2)$. In *Approximation, randomization, and combinatorial optimization*, volume 2764 of *Lecture Notes in Comput. Sci.*, pages 188–199. Springer, Berlin, 2003.

[4] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. Twenty-first Symposium on the Theory of Computing (Seattle, WA, 1989).

[5] E. Ben-Sasson, M. Sudan, S. Vadhan, and A. Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, pages 612–621 (electronic), New York, 2003. ACM.

[6] C. Bertram-Kretzberg and H. Lefmann. Mod_p-tests, almost independence and small probability spaces. *Random Struct. Algorithms*, 16(4):293–313, 2000.

[7] A. Bogdanov. Pseudorandom generators for low degree polynomials. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 21–30, New York, 2005. ACM.

[8] O. Goldreich. *Modern cryptography, probabilistic proofs and pseudorandomness*, volume 17 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1999.

[9] W. T. Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998.

[10] W. T. Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001.

[11] B. Green and T. Tao. An inverse theorem for the gowers u^3 norm, 2005. arXiv.org:math/0503014.

[12] R. Lidl and H. Niederreiter. *Finite fields*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, second edition, 1997. With a foreword by P. M. Cohn.

[13] S. Lovett. Pseudorandom generators for low degree polynomials, 2007. Manuscript.

[14] M. Luby, B. Velickovic, and A. Wigderson. Deterministic approximate counting of depth-2 circuits. In *Proceedings of the 2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993.

[15] J. Naor and M. Naor. Small-bias probability spaces: efficient constructions and applications. In *Proceedings of the 22nd Annual ACM Symposium on the Theory of Computing*, pages 213–223, 1990.

[16] A. Samorodnitsky. Low-degree tests at large distances. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, CA USA*, 2007.

[17] A. Samorodnitsky and L. Trevisan. Gowers uniformity, influence of variables, and PCPs. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing, Seattle, WA, USA*, pages 11–20, 2006.

[18] V. Shoup. New algorithms for finding irreducible polynomials over finite fields. *Math. Comp.*, 54(189):435–447, 1990.

[19] E. Viola. New correlation bounds for gf(2) polynomials using gowers uniformity. *Electronic Colloquium on Computational Complexity*, Technical Report TR06-097, 2006. `http://www.eccc.uni-trier.de/eccc`.

[20] E. Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007.

[21] E. Viola and A. Wigderson. Norms, xor lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the 22nd Annual Conference on Computational Complexity*. IEEE, June 13–16 2007.