Pseudorandom bits for polynomials

Andrej Bogdanov* Emanuele Viola †

March 1, 2009

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 finite 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 "inverse conjecture for the Gowers norm," 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

†Supported by NSF grant CCF-0845003. This work was partially done when the author was a postdoctoral fellow at the Institute for Advanced Study supported by NSF grant CCR-0324906, and a postdoctoral fellow at Columbia University supported by NSF award CCF-0347282 and NSF award CCF-0523664. Email: viola@ccs.neu.edu

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 [Gol99] 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 [NN93] who gave a generator with seed length $s = O(\log n)$ (for error $\epsilon = 1/n$), which is optimal up to constant factors (cf. [AGHP92]). This generator, also known as small-bias generator, has been one of the most celebrated results in pseudorandomness, with applications ranging from derandomization [NN93], to PCP's [BSSVW03], and to lower bounds [BNS92, VW08], just to name a few (cf. references in [BSSVW03]).

Subsequently, Luby, Veličković, and Wigderson (Theorem 2 in [LVW93]; cf. [Vio07]) 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 [LVW93], whereas $n = 2^{\Omega(s)}$ in [NN93].) Bogdanov [Bog05] 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 [LVW93], even for the case of quadratic polynomials.

1.1 Our results

In this work we construct the following generators.

Theorem 1. *Over any finite 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 $s =$*

[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] [Bog05] also gives generators over small fields, but in this case the seed length is worse than what can be obtained from [LVW93].

$O(\log_{|\mathbb{F}|} n) + f(\epsilon, \mathbb{F})$, 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, cf. Appendix A.

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

Theorem 2. *Assume that the "d vs. $d-1$ inverse conjecture for the Gowers norm" (Conjecture 21) 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. $d-1$ inverse conjecture for the Gowers norm may be significantly easier to prove than the general one (cf. discussion after Conjecture 21). We also point out that a strengthening of known "inverse results" [GT08, Sam07] 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 [NN93, AGHP92, ABN+92].

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 [Gow98, Gow01] and independently by Alon et al. [AKK+03], and has found a wide variety of applications, ranging from arithmetic combinatorics [Gow98, Gow01, GT08], property testing [AKK+03], PCP's [ST06, Sam07], and lower bounds [Vio06, VW08]. 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, our results apply over arbitrary finite fields as we point out later.

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

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

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

$$\text{Correlation}(f, q) := \left|\Pr_{X \in \mathbb{F}_2^n}[f(X) = q(X)] - \Pr_{X \in \mathbb{F}_2^n}[f(X) \neq q(X)]\right|.$$

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. [Gow98, Gow01, GT08, VW08]); the converse is known to hold for $d = 2$ with polynomial slackness in the parameters (see, e.g., [GT08, Sam07]), for $d = 3$ with exponential slackness in the parameters [GT08, Sam07], and is conjectured to hold for any fixed d (see [GT08, Section 13] and [Sam07]). This latter conjecture is usually referred to as "the inverse conjecture for the Gowers norm." In this discussion we ignore both the status and the quantitative aspect of ($\star$) and we proceed with the intuition behind our approach. We mention however that research subsequent to this paper proves the conjecture true in certain cases, and false in others; see the discussion at the end of this section.

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:

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

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:

$$\text{Bias}(p(X)) \leq U_d(p)^{1/2^d}. \tag{2}$$

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

$$\text{Bias}(p(W)) \leq U_d(p)^{1/2^d} + \epsilon, \tag{3}$$

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

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

where the approximation holds because q_p is block-linear and each of the W_i fools linear tests (here the Y_i's are independent as well, and we use 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

   $$U_2(p) = U_2(p \circ A) = U_2(x_1x_2 + x_3x_4 + \ldots + x_{2t-1}x_{2t} + \ell(x)) = 2^{-2t}.$$

   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.

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. It is natural to ask whether the seed length of our generators can be improved by using fewer than d independent copies of the linear generator. In Appendix A we show this is not possible: The sum of $d-1$ copies of generators that fool linear polynomials in general does not fool degree d polynomials. Appendix B covers some details related to our definition of "pseudorandom." We work over prime fields for most of this paper, and in Appendix C we point out how one can obtain generators over non prime fields as well.

Subsequent developments. Our work was followed by a flurry of research, which we now overview chronologically. Shachar Lovett [Lov08] shows unconditionally that the sum of $2^{O(d)}$ generators that $\epsilon^{2^{O(d)}}$-fool linear functions fools degree-d polynomials with error ϵ. Green and Tao [GT07] prove the d vs. $d-1$ inverse conjecture for the Gowers norm *when the field size $|\mathbb{F}|$ is bigger than the degree d of the polynomial*. In such cases ($|\mathbb{F}| > d$), the combination of [GT07] with this work shows that the sum of d generators that ϵ'-fool linear functions fools degree-d polynomials with error ϵ, where ϵ' depends on $d, |\mathbb{F}|$, and ϵ, but not on n. The proof by Green and Tao [GT07] builds on our Lemma 24. On the negative side, Green and Tao [GT07], and independently Lovett, Meshulam, and Samorodnitsky [LMS08], show that the d vs. $d-1$ inverse conjecture for the Gowers norm is *false* when the field size is much smaller than the degree of the polynomial (which in particular falsifies the more general inverse conjecture for the Gowers norm [GT08, Sam07]). This falsity prevents the analysis in this work to go through for large degree and small fields such as $\mathbb{F}_2 = \{0,1\}$. Via a different analysis, Viola [Vio08] shows that the sum of d generators that ϵ'-fool linear functions ϵ-fools degree-d polynomials *over any field* $\mathbb{F}$, where ϵ' depends only on $d, |\mathbb{F}|$, and ϵ, but not on n. Over $\mathbb{F}_2$, one can take $\epsilon' := (\epsilon/16)^{2^{d-1}}$.

2 Preliminaries

In this paper we obtain generators over arbitrary finite fields. We actually work over prime fields for most of the paper, and then we point out in Appendix C how the results extend to non-prime fields. Although all our main ideas are already present in the results for the fundamental case $\mathbb{F}_2 = \{0, 1\}$, we obtain results about arbitrary prime fields at essentially no additional cost, and thus we present our results in this generality. For this goal, it is useful to introduce the following notation.

Notation 3. *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*

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

to get rid of a few brackets.

The notion of pseudorandomness that we use is the following.

Definition 4 (Pseudorandomness)**.** *Let $\mathcal{F}$ be a family of functions from $\mathbb{F}^n$ to $\mathbb{F}$. We say that a distribution W on $\mathbb{F}^n$* fools *$\mathcal{F}$ with error ϵ if for every $f \in \mathcal{F}$ we have*

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

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

Here, we will be mostly interested in the case where $\mathcal{F}$ is the family of degree-d polynomials over $\mathbb{F}$.

Remark 5 (On Definition 4)**.** *We point out that pseudorandomness is often defined in terms of* statistical distance*. However, the algebraic Definition 4 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 we formally prove this in Appendix B.*

The basic building block of our construction is a generator for degree-1 polynomials. This generator was first obtained for $\mathbb{F}_2$ in [NN93]. Then [AGHP92] 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 6 ([NN93], Proposition 4.1 in [BKL00])**.** *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 [BKL00] is a straightforward extension of the "powering" construction in [AGHP92] 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 [Sho90]. 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 $f(x)$ is a polynomial of degree d, all its directional derivatives are polynomials of degree at most $d-1$ (in the variable x). We can take multiple derivatives too: We write $\mathrm{D}_{y_1,\ldots,y_k} f(x)$ for the function

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

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 7. *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 [Gow01, GT08] (a proof also appears in [ST06]).

Definition 8 (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*

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

where X and the Y_i's are independent and uniformly distributed over $\mathbb{F}^n$.

[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., [GT08].

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 [Gow01, GT08] (cf. [VW08, Lemma 2.3]): For every degree-d polynomial $p : \mathbb{F}^n \to \mathbb{F}$ over a prime field $\mathbb{F}$,

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

The generalization that we prove is the following.

Lemma 9. *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*

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

Before discussing the proof of Lemma 9 we make some remarks.

Remark 10. *We observe the following. (1) Lemma 9 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 9 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 $|\mathrm{E}_X\, e[p(X)]|$ and $|\mathrm{E}_{W_1,\ldots,W_d}\, e[p(W_1 + \cdots + W_d)]|$ are small, and so is their difference by the triangle inequality.*

We now discuss the proof of Lemma 9. For the proof we need some claims.

Definition 11. *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 12. *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].

Claim 13. *For every function $f : \mathbb{F}^n \to \mathbb{F}$ and every distribution $\mathcal{D}$ over $\mathbb{F}^n$,*

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

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 9.

[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

Proof of Lemma 9 from Claims 12 and 13. Let $q_p(y_1, \ldots, y_d) := \mathrm{D}_{y_1,\ldots,y_d}\, p(x)$.

$$\begin{aligned}
&\left|\mathrm{E}_{W_1,\ldots,W_d}\, e\left[p(W_1 + \cdots + W_d)\right]\right|^{2^d} \\
&\quad\le \mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}}\, e\left[\mathrm{D}_{W'_1 - W_1,\ldots,W'_d - W_d}\, p(W_1 + \cdots + W_d)\right] && \text{by Claim 13} \\
&\quad= \mathrm{E}_{\substack{W_1,\ldots,W_d \\ W'_1,\ldots,W'_d}}\, e\left[q_p(W'_1 - W_1, \ldots, W'_d - W_d)\right] \\
&\quad\le \mathrm{E}_{Y_1,\ldots,Y_d \in \mathbb{F}^n}\, e\left[q_p(Y_1, \ldots, Y_d)\right] + \epsilon \cdot d && (5) \\
&\quad= U_d\left(p\right) + \epsilon \cdot d,
\end{aligned}$$

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 12, 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

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

Proof of Claim 12. Let us first assume that p is a monomial of degree at most d with coefficient 1. If p has degree strictly smaller than d then $\mathrm{D}_{y_1,\ldots,y_d}\, p(x) = 0$ and the conclusion holds trivially. If p has degree exactly d, then $p(x)$ is of the form $x_{j_1} \cdots x_{j_d}$ (with possible repetitions among the indices). A calculation shows that $D_{y_1,\ldots,y_d} p(x) = \mathrm{per}(Y)$, where per is the permanent polynomial of a $d \times d$ matrix and Y is the matrix with entries $Y_{ij} = y_{i,j_i}$. The multilinearity of $D_{y_1,\ldots,y_d} p(x)$ then follows from the fact that the permanent polynomial is block-linear in the rows of the matrix. In general, a degree d polynomial p is a linear combination of monomials with coefficient 1. Taking derivatives is a linear operation, so $\mathrm{D}_{y_1,\ldots,y_d}\, p(x)$ is a linear combination of block-linear polynomials with blocks $y_1, \ldots, y_d$. Block-linear polynomials (with the same block structure) are closed under linear operations, so q_p is block-linear. (See Fact 16 for examples.) $\square$

Proof of Claim 13. We proceed by induction on d. When $d = 1$ we have

$$\begin{aligned}
\left|\mathrm{E}_{W_1}\, e\left[f(W_1)\right]\right|^2 &= \mathrm{E}_{W'_1}\, e\left[f(W'_1)\right] \cdot \overline{\mathrm{E}_{W_1}\, e\left[f(W_1)\right]} \\
&= \mathrm{E}_{W_1, W'_1}\, e\left[f(W'_1) - f(W_1)\right] = \mathrm{E}_{W_1, W'_1}\, e\left[\mathrm{D}_{W'_1 - W_1}\, f(W_1)\right].
\end{aligned}$$

For $d > 1$ we have, using several times the fact that $|\mathrm{E}_Z[Z]|^2 \le \mathrm{E}_Z[|Z|^2]$ for any complex random variable Z,

$$
\begin{aligned}
&\left|\mathrm{E}_{W_1,\ldots,W_d} e\left[f(W_1+\cdots+W_d)\right]\right|^{2^d} \\
&\quad\le \mathrm{E}_{W_1,\ldots,W_{d-1}}\left[\left|\mathrm{E}_{W_d} e\left[f(W_1+\cdots+W_d)\right]\right|^2\right]^{2^{d-1}} \\
&\quad= \mathrm{E}_{W_1,\ldots,W_{d-1}}\left[\mathrm{E}_{W_d,W'_d} e\left[\mathrm{D}_{W'_d-W_d} f(W_1+\cdots+W_d)\right]\right]^{2^{d-1}} && \text{by the case } d=1 \\
&\quad\le \mathrm{E}_{W_d,W'_d}\left[\left|\mathrm{E}_{W_1,\ldots,W_{d-1}} e\left[\mathrm{D}_{W'_d-W_d} f(W_1+\cdots+W_d)\right]\right|^{2^{d-1}}\right] \\
&\quad\le \mathrm{E}_{W_d,W'_d}\left[\mathrm{E}_{\substack{W_1,\ldots,W_{d-1}\\W'_1,\ldots,W'_{d-1}}} e\left[\mathrm{D}_{W'_1-W_1,\ldots,W'_d-W_d} f(W_1+\cdots+W_d)\right]\right] && \text{by inductive hypothesis} \\
&\quad= \mathrm{E}_{\substack{W_1,\ldots,W_d\\W'_1,\ldots,W'_d}} e\left[\mathrm{D}_{W'_1-W_1,\ldots,W'_d-W_d} f(W_1+\cdots+W_d)\right]. && \square
\end{aligned}
$$

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

In this section we prove the following lemma.

Lemma 14. *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*

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

Note that Lemma 14 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 15. *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 sketched at the end of this section.

Fact 16 (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* $(f \oplus f') : \mathbb{F}^{n+n'} \to \mathbb{F}$ *be the function* $(f \oplus f')(x, x') := f(x) + f'(x')$*. Then* $U_k(f \oplus f') = U_k(f) \cdot U_k(f')$*.*
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*
$$U_2(f) = \mathrm{E}_{Y_1, Y_2 \in \mathbb{F}}\, e[2 \cdot a \cdot Y_1 \cdot Y_2] = 1/|\mathbb{F}|.$$
5. *Let* $f : \mathbb{F}_2^2 \to \mathbb{F}_2$ *be* $f(x_1, x_2) := x_1 \cdot x_2$*. Then*
$$U_2(f) = \mathrm{E}_{Y_{11}, Y_{12}, Y_{21}, Y_{22} \in \mathbb{F}_2}\, e[Y_{11} \cdot Y_{22} + Y_{21} \cdot Y_{12}] = 1/4.$$

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 17 (Theorems 6.21 and 6.30 in [LN97])**.** *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* $|\mathbb{F}| = 2$ *then* $(p \circ A)(x) = \sum_{1 \leq i \leq \lfloor n/2 \rfloor} c_i \cdot x_{2i-1} \cdot x_{2i} + \ell(x)$*,*
2. *If* $|\mathbb{F}|$ *is odd then* $(p \circ A)(x) = \sum_{1 \leq i \leq n} c_i \cdot x_i^2 + \ell(x)$*.*

Armed with the above results, we now present the proof of Lemma 15.

Proof of Lemma 15. 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 proved analogously, using Item 5 in Fact 16. By Lemma 17, 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: $(p \circ A)(x) = \sum_{1 \leq i \leq n} c_i \cdot x_i^2 + \ell(x)$. 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:

$$\begin{aligned}
U_2(p) = U_2(p \circ A) &= U_2\left(\sum\nolimits_{1 \leq i \leq n} c_i \cdot x_i^2 + \ell(x)\right) && \text{by Fact 16.2} \\
&= U_2\left(\sum\nolimits_{1 \leq i \leq n} c_i \cdot x_i^2\right) = U_2\left(\sum\nolimits_{1 \leq i \leq s} c_i \cdot x_i^2\right) && \text{by Fact 16.3} \\
&= \prod\nolimits_{1 \leq i \leq s} U_2\left(c_i \cdot x_i^2\right) && \text{by Fact 16.1} \\
&= 1/|\mathbb{F}|^s && \text{by Fact 16.4.}
\end{aligned}$$

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 16. 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

$$U_k\left(f \circ A\right) = \mathrm{E}_{Y_1,Y_2,\ldots,Y_k,X\in\mathbb{F}^n}\, e\left[\sum_{S\subseteq[k]}(-1)^{k-|S|} f\circ A(X+\sum_{i\in S}Y_i)\right]$$

$$= \mathrm{E}\, e\left[\sum_{S\subseteq[k]}(-1)^{k-|S|} f(AX+\sum_{i\in S}AY_i)\right] = \mathrm{E}\, e\left[\sum_{S\subseteq[k]}(-1)^{k-|S|} f(X+\sum_{i\in S}Y_i)\right] = U_k\left(f\right).$$

(All expectations are over the uniform distribution.) 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 15 proves the main Lemma 14 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 $h(x)=\sum_{a\in\mathbb{F}^k}\hat{h}_a e_a(x)$, where e_a is the function $e_a(x):=e(a_1x_1+\cdots+a_kx_k)$ and the *Fourier coefficients* $\hat{h}_a$ are given by $\hat{h}_a = E_x\big[h(x)\cdot\overline{e_a(x)}\big]$. The Fourier coefficients satisfy Parseval's identity

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

We state the following lemma in higher 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 18. *Let $\mathcal{F}$ be any class of functions $\mathbb{F}^n\to\mathbb{F}$ that forms a linear space over $\mathbb{F}$. Suppose that W fools $\mathcal{F}$ with error ϵ. Then for every function $h:\mathbb{F}^k\to\mathbb{C}$ and every collection of functions $f_1,\ldots,f_k\in\mathcal{F}$,*

$$\left|\mathrm{E}_{X\sim\mathbb{F}^n}[h(f_1(X),\ldots,f_k(X))] - \mathrm{E}_W[h(f_1(W),\ldots,f_k(W))]\right| \leq \epsilon\cdot L(h),$$

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

Before proving Lemma 18, let us see how it can be used to prove the main Lemma 14 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 19. *Let $h : \mathbb{F}^k \to \mathbb{C}$ be a function such that $|h(x)| = 1$. Then $L(h) \le |\mathbb{F}|^{k/2}$.*

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

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

Proof of Lemma 14 assuming Lemma 18. By Lemma 15 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 18 and Proposition 19 we have that

$$|\mathrm{E}_X \, e[p(X)] - \mathrm{E}_W \, e[p(W)]| \le |\mathbb{F}|^{t/2} \cdot \epsilon = O\left(\frac{\epsilon}{U_2(p)}\right). \qquad \square$$

It now only remains to prove Lemma 18.

Proof of Lemma 18. For every distribution Y on $\mathbb{F}^k$, we have that

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

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

$$\mathrm{E}_Y[e_a(Y)] = \mathrm{E}_Y \, e[a_1 Y_1 + \cdots + a_k Y_k] = \mathrm{E}_Z \, e[a_1 f_1(Z) + \cdots + a_k f_k(Z)].$$

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

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

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

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

5 Fooling quadratic polynomials

Theorem 20 (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*

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

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 $U_2(p)$. Let $\tau := 1/n$ (note that the error of the W's is τ^c).

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

$$\begin{aligned} |\mathrm{E}_X\, e[p(X)] - \mathrm{E}_W\, e[p(W)]| &\leq |\mathrm{E}_X\, e[p(X)]| + |\mathrm{E}_W\, e[p(W)]| && \text{by the triangle inequality} \\ &\leq (\tau^{c/2})^{1/4} + (\tau^{c/2} + 2 \cdot \tau^c)^{1/4} && \text{by using Lemma 9 twice,} \end{aligned}$$

which, for sufficiently large c, proves the lemma in this case.

Case $U_2(p) \geq \tau^{c/2}$. In this case Lemma 14 gives

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

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 6). □

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

In this section we prove a lemma which is similar to Lemma 14 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 "inverse conjecture for the Gowers norm," which we now describe.

Conjecture 21 (d vs. $d-1$ inverse conjecture for the Gowers norm 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:*

$$U_d(p) \geq \tau \Rightarrow \max_{\substack{q:\, \mathbb{F}^n \to \mathbb{F} \\ \text{of degree } d-1}} |\mathrm{E}_{X \in \mathbb{F}^n}\, e[p(x) - q(x)]| \geq IG^d_{\mathbb{F}}(\tau).$$

Remark 22. *Conjecture 21 is usually stated for* arbitrary *functions p [GT08, Sam07]. However, for the results in this work the special case where p has degree d is sufficient. We call this the d vs. $d-1$ inverse conjecture for the Gowers norm. We stress that subsequent to our work there has been progress on these conjectures; see the end of Section 1.*

The following lemma says that, if we believe the d vs. $d-1$ inverse conjecture for the Gowers norm 21, 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 23. *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 $U_d(p) \geq \tau$ and that Conjecture 21 holds for $\mathbb{F}$ and d. Then*

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

where $K \leq (|\mathbb{F}|/IG(\tau))^c$ for a universal constant $c > 0$.

To prove Lemma 23, we would like a stronger statement than what is given to us by the inverse conjecture for the Gowers norm. 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 23 using Lemma 18.

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 24. *Let $\gamma > 0$ and $f, g : \mathbb{F}^n \to \mathbb{F}$, for a prime field $\mathbb{F}$. Suppose that*

$$\left|\mathrm{E}_{x \in \mathbb{F}^n}\, e\left[f(x) - g(x)\right]\right| \geq \delta.$$

Then there is an integer t and a function $B \colon \mathbb{F}^t \to \mathbb{F}$ such that for every $x \in \mathbb{F}^n$

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

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

Lemma 24 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 to be required over larger fields.

Intuition for the proof of Lemma 24. 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 $p(s) := \Pr_a[g(a) - f(a) = s]$).

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

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

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 $p_{-r'}(s) = \Pr_a[v(a) = s]$ 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 (using the fact that $|\mathbb{F}|$ is prime), which can be shown to contradict our assumption that $|\mathrm{E}_{x \in \mathbb{F}^n} e[f(x) - g(x)]| \geq \delta$ (jumping ahead, via Claim 33).

This suggests a natural approach to the decoding problem: On input x, compute the probabilities $\Pr_a[v(a) = s]$ 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.

The statistical distance between two distributions p, q is denoted by $\mathrm{SD}(p, q)$, and in the following proof we use some standard facts about this quantity. A discussion of statistical distance together with proofs of the relevant facts appears in Section B.

Proof of Lemma 24. First we present the algorithm for "decoding" $f(x)$; then we prove its correctness. Let $p(s) := \Pr_a[g(a) - f(a) = s]$.

Algorithm for computing $f(x)$ given p and oracle access to v:
On input x, do the following.

1. Query v at t random locations $a_1, \ldots, a_t$.
2. Compute the empirical frequency of each $y \in \mathbb{F}$ among the answers to the queries: Namely, compute the distribution q on $\mathbb{F}$ where
$$q(y) := \frac{1}{t} \cdot \left|\{i : v(a_i) = y\}\right|.$$
3. Return the value $r' \in \mathbb{F}$ that minimizes $\mathrm{SD}(q, p_{-r'})$.

We now analyze the above algorithm. We must show that, for t as in the statement of the lemma, with probability $1 - \gamma$ we have

$$\mathrm{SD}(p_{-r}, q) < \mathrm{SD}(p_{-r'}, q)$$

for every $r' \neq r$ (recall that $r = f(x)$). To show this, we make two claims.

Claim 25. *For every $r \neq r'$, $\mathrm{SD}(p_{-r}, p_{-r'}) > \delta/|\mathbb{F}|$.*

Claim 26. *With probability $1 - \gamma$, $\mathrm{SD}(p_{-r}, q) \leq \delta/2|\mathbb{F}|$.*

From the above two claims the lemma follows easily. Specifically, with probability $1 - \gamma$ for every $r' \neq r$,

$$\begin{aligned}
\mathrm{SD}(p_{-r}, q) &\leq \delta/2|\mathbb{F}| && \text{by Claim 26} \\
&< \mathrm{SD}(p_{-r}, p_{-r'}) - \delta/2|\mathbb{F}| && \text{by Claim 25} \\
&\leq \big(\mathrm{SD}(p_{-r}, q) + \mathrm{SD}(p_{-r'}, q)\big) - \delta/2|\mathbb{F}| && \text{by the triangle inequality} \\
&\leq \mathrm{SD}(p_{-r'}, q) && \text{by Claim 26.}
\end{aligned}$$

□

To complete the proof it only remains to prove the above two claims.

Proof of Claim 25. Let $i := r' - r \neq 0$. Then $\mathrm{SD}(p_{-r}, p_{-r'}) = \mathrm{SD}(p, p_{-i})$. We will show that

$$\mathrm{SD}(p, p_{-i}) \geq \frac{2}{|\mathbb{F}| - 1} \cdot \mathrm{SD}(p, u) \tag{9}$$

where u is the uniform distribution on $\mathbb{F}$. The assumption $|\mathrm{E}_x\, e\,[f(x) - g(x)]| \geq \delta$ implies that $\mathrm{SD}(p, u) \geq \delta/2$ (using Claim 33) which will conclude the proof.

We prove equation (9) by the following calculation, which repeatedly uses the triangle inequality $|a+b| \leq |a| + |b|$. First, for every $y \in \mathbb{F}$,

$$\begin{aligned}
\big|p(y) - u(y)\big| &= \left|\frac{1}{|\mathbb{F}|}\sum_{z\in\mathbb{F}}(p(y) - p(z))\right| \\
&\leq \frac{1}{|\mathbb{F}|}\sum_{z\neq y}\big|p(y) - p(z)\big| \\
&= \frac{1}{|\mathbb{F}|}\sum_{t\neq 0}\big|p(y) - p(y - t\cdot i)\big| \\
&\leq \frac{1}{|\mathbb{F}|}\sum_{t\neq 0}\sum_{s=0}^{t-1}\big|p(y - s\cdot i) - p(y - (s+1)i)\big| \\
&= \frac{1}{|\mathbb{F}|}\sum_{t\neq 0}\sum_{s=0}^{t-1}\big|p(y - s\cdot i) - p_{-i}(y - s\cdot i)\big|.
\end{aligned}$$

It follows that

$$\begin{aligned}
2\cdot \textsc{sd}(p,u) &= \sum_{y\in\mathbb{F}}|p(y) - u(y)| \\
&\leq \frac{1}{|\mathbb{F}|}\sum_{y\in\mathbb{F}}\sum_{t\neq 0}\sum_{s=0}^{t-1}\big|p(y - s\cdot i) - p_{-i}(y - s\cdot i)\big| \\
&= \frac{1}{|\mathbb{F}|}\sum_{t\neq 0}\sum_{s=0}^{t-1}\sum_{y\in\mathbb{F}}\big|p(y - s\cdot i) - p_{-i}(y - s\cdot i)\big| \\
&= \frac{1}{|\mathbb{F}|}\sum_{t\neq 0}\sum_{s=0}^{t-1}\sum_{y\in\mathbb{F}}\big|p(y) - p_{-i}(y)\big| \\
&= \frac{1}{|\mathbb{F}|}\frac{(|\mathbb{F}|-1)|\mathbb{F}|}{2}\sum_{y\in\mathbb{F}}\big|p(y) - p_{-i}(y)\big| \\
&= (|\mathbb{F}|-1)\cdot \textsc{sd}(p, p_{-i}).
\end{aligned}$$

□

Proof of Claim 26. Set $\eta := \delta/(2\cdot|\mathbb{F}|)$. By the definition of statistical distance we have that

$$\begin{aligned}
\Pr_{a_1,\ldots,a_t}[\textsc{sd}(p_{-r}, q) > \eta] &= \Pr_{a_1,\ldots,a_t}\big[\exists S \subseteq \mathbb{F} : \Pr_X[X\in S] - \Pr_Y[Y\in S] > \eta\big] \\
&\leq \sum_{S\subseteq\mathbb{F}}\Pr_{a_1,\ldots,a_t}\big[\Pr_Y[Y\in S] < \Pr_X[X\in S] - \eta\big],
\end{aligned}$$

where X and Y are random variables that denote a sample from p_{-r} and q, respectively. We will bound each term in the summation separately using the Chernoff bound. For this we

fix $a_1, \ldots, a_t$ and write

$$\Pr_Y[Y \in S] = \sum_{y \in S} q(y) = \sum_{y \in S} \frac{|\{i : v(a_i) = y\}|}{t} = \frac{|\{i : v(a_i) \in S\}|}{t} = \frac{Z_1 + \cdots + Z_t}{t}$$

where Z_i is an indicator variable for the event $v(a_i) \in S$.

If we now view Z_i as random variables over the choice $a_1, \ldots, a_t$, we have that the Z_i are i.i.d with mean

$$\mathrm{E}_{a_i}[Z_i] = \Pr_{a_i}[v(a_i) \in S] = \Pr_X[X \in S]$$

since each $v(a_i)$ is sampled from the distribution p_{-r}. Therefore

$$\begin{aligned}\Pr_{a_1,\ldots,a_t}\big[\Pr_Y[Y \in S] < \Pr_X[X \in S] - \eta\big] \\ = \Pr_{a_1,\ldots,a_t}\big[Z_1 + \cdots + Z_t < \mathrm{E}[Z_1 + \cdots + Z_t] - t\eta\big] \le e^{-2\eta^2 t},\end{aligned}$$

where the last inequality is a form of the Chernoff bound (e.g. [DP05, Theorem 1.1]). It follows that for $t := (|\mathbb{F}|/\delta)^c(1 + \log 1/\gamma)$, where c is a sufficiently large absolute constant, we have

$$\Pr_{a_1,\ldots,a_t}\big[\mathrm{SD}(p_{-r}, q) \le \eta\big] \ge 1 - 2^{|\mathbb{F}|} \cdot e^{-2\eta^2 t} \ge 1 - \gamma. \qquad \square$$

6.2 Proof of Lemma 23

We now prove Lemma 23.

Proof of Lemma 23. By assumption and Conjecture 21, there is a polynomial q of degree $d-1$ such that

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

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

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

be the function in Lemma 24 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

$$\begin{aligned}\big|\mathrm{E}_Y\, e\,[p(Y)] - \mathrm{E}_{A,Y}\, e\,[h(A, Y)]\big| &\le \mathrm{E}_Y\, \mathrm{E}_A\big[|e(p(Y)) - e(h(A, Y))|\big] \\ &\le \mathrm{E}_Y\big[2 \cdot \Pr_A[p(Y) \ne h(A, Y)]\big] \\ &< 2 \cdot \gamma, \qquad (10)\end{aligned}$$

where the last inequality follows from Lemma 24.

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 fools degree $d-1$ polynomials with error ϵ, by Lemma 18 and Proposition 19 we have that

$$\text{For every } a,\ \left|\mathrm{E}_X\, e\left[h(a,X)\right] - \mathrm{E}_W\, e\left[h(a,W)\right]\right| \le \epsilon \cdot |\mathbb{F}|^{t/2} \tag{11}$$

and therefore

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

where the second inequality follows from (10) and the third one is inequality (11). 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 24 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 inverse conjecture for the Gowers norm, 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 inverse theorems [GT08, Sam07] 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 27. *Let a field $\mathbb{F}$ and a degree d be given. Assume Conjecture 21 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 $d-1$ independent random variables that fool linear functions on n inputs with error ϵ_1 fools degree $d-1$ polynomials on n inputs with error ϵ_{d-1}.

Let $W := W_1 + \cdots + W_d$ and p be an arbitrary degree-d polynomial. We proceed analogously to Theorem 20, 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 9 we have that

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

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

Case $U_d(p) \ge \tau$. From Lemma 23 we have

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

where $K \le (|\mathbb{F}|/IG(\tau))^{O(1)}$. 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 $\exp(K)\cdot\epsilon_{d-1}^{1/K} \le \epsilon_d$ 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 inverse conjecture for the Gowers norm was proved by Green and Tao [GT08] and Samorodnitsky [Sam07], though with exponential slackness in the parameters.

Theorem 28 ([GT08, Sam07])**.** *Fix a prime field $\mathbb{F}$.*[6] *For every n and every function $f : \mathbb{F}^n \to \mathbb{F}$ the following is true:*

$$U_3(p) \ge \tau \Rightarrow \max_{\substack{q:\,\mathbb{F}^n\to\mathbb{F} \\ \text{of degree } 2}} \left|\mathrm{E}_{X\in\mathbb{F}^n}\, e\left[p(x) - q(x)\right]\right| \ge IG^3_{\mathbb{F}}(\tau),$$

where

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

for an absolute constant $C > 0$.

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

Combining our approach with Theorem 28 we obtain the following unconditional generator for cubic polynomials.

Theorem 29. *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 27. Let us think of being given $\epsilon = \epsilon_3$. In the proof of Theorem 27 the total error is dominated by the error in the case of large norm. In this case, the error is $\exp(K_1) \cdot \epsilon_2^{1/K_1}$, where $K_1 \leq (|\mathbb{F}|/IG(\tau))^{O(1)}$ and ϵ_2 is such that $W_1 + W_2 + W_3$ ϵ_2-fools quadratic polynomials. Since we are fixing the field, in the expression for K_1 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 28 we obtain that

$$K_1 \leq K_2 \cdot \exp\left(1/\epsilon_3^{K_2}\right)$$

for a universal constant K_2. Consequently, the error is at most ϵ_3 if

$$\exp(K_2 \cdot \exp\left(1/\epsilon_3^{K_2}\right)) \cdot \epsilon_2^{1/\left(K_2 \cdot \exp\left(1/\epsilon_3^{K_2}\right)\right)} \leq \epsilon_3.$$

The above inequality is satisfied for $\epsilon_2 = \exp\left(-\exp\left(1/\epsilon_3^{K_3}\right)\right)$ for a sufficiently large universal constant K_3. The required bound on ϵ_2 follows by assumption (for sufficiently large c) using Theorem 20. □

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 [GT08] and [Sam07]. We also thank the anonymous referees for their feedback.

References

[ABEK08] Noga Alon, Ido Ben-Eliezer, and Michael Krivelevich. Small sample spaces cannot fool low degree polynomials. In *12th Workshop on Randomization and Computation (RANDOM)*, pages 266–275. Springer, 2008. 25

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron 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, 25, 26

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. 1, 2, 6, 7, 25

[AKK+03] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. Testing low-degree polynomials over GF(2). In *7th International Workshop on Randomization and Approximation Techniques in Computer Science (RANDOM)*, volume 2764 of *Lecture Notes in Computer Science*, pages 188–199. Springer, 2003. 2

[BKL00] Claudia Bertram-Kretzberg and Hanno Lefmann. Mod_p-tests, almost independence and small probability spaces. *Random Struct. Algorithms*, 16(4):293–313, 2000. 6, 7

[BNS92] László Babai, Noam Nisan, and Márió Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. System Sci.*, 45(2):204–232, 1992. 1

[Bog05] Andrej Bogdanov. Pseudorandom generators for low degree polynomials. In *37th Annual Symposium on Theory of Computing (STOC)*, pages 21–30. ACM, 2005. 1

[BSSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short PCPs via epsilon-biased sets. In *35th Annual Symposium on Theory of Computing (STOC)*, pages 612–621. ACM, 2003. 1

[DP05] Devdatt Dubhashi and Alessandro Panconesi. Concentration of measure for the analysis of randomised algorithms, 2005. http://www.dsi.uniroma1.it/~ale/papers.html. 19

[Gol99] Oded Goldreich. *Modern cryptography, probabilistic proofs and pseudorandomness*, volume 17 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1999. 1

[Gow98] Timothy Gowers. A new proof of Szemerédi's theorem for arithmetic progressions of length four. *Geom. Funct. Anal.*, 8(3):529–551, 1998. 2, 3

[Gow01] Timothy Gowers. A new proof of Szemerédi's theorem. *Geom. Funct. Anal.*, 11(3):465–588, 2001. 2, 3, 7, 8

[GT07] Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms, 2007. arXiv:0711.3191v1. 5

[GT08] Ben Green and Terence Tao. An inverse theorem for the Gowers $U^3(G)$ norm. *Proceedings of the Edinburgh Mathematical Society (Series 2)*, 51(01):73–153, 2008. 2, 3, 5, 7, 8, 14, 20, 21, 22

[LMS08] Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. In *40th Annual Symposium on the Theory of Computing (STOC)*, pages 547–556. ACM, 2008. 5

[LN97] Rudolf Lidl and Harald Niederreiter. *Finite fields*, volume 20 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, second edition, 1997. 11

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In *40th Annual Symposium on the Theory of Computing (STOC)*, pages 557–562. ACM, 2008. 5

[LVW93] Michael Luby, Boban Veličković, and Avi Wigderson. Deterministic approximate counting of depth-2 circuits. In *2nd Israeli Symposium on Theoretical Computer Science (ISTCS)*, pages 18–24, 1993. 1

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM J. Comput.*, 22(4):838–856, 1993. 1, 2, 6, 25

[Sam07] Alex Samorodnitsky. Low-degree tests at large distances. In *39th Annual Symposium on Theory of Computing (STOC)*, pages 506–515. ACM, 2007. 2, 3, 5, 14, 20, 21, 22

[Sho90] Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. *Math. Comp.*, 54(189):435–447, 1990. 7

[ST06] Alex Samorodnitsky and Luca Trevisan. Gowers uniformity, influence of variables, and PCPs. In *38th Symposium on Theory of Computing (STOC)*, pages 11–20. ACM, 2006. 2, 7

[Vio06] Emanuele Viola. New correlation bounds for GF(2) polynomials using Gowers uniformity. *Electronic Colloquium on Computational Complexity*, Technical Report TR06-097, 2006. www.eccc.uni-trier.de/. 2

[Vio07] Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007. 1

[Vio08] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *23nd Annual Conference on Computational Complexity*. IEEE, June 23–26 2008. http://www.ccs.neu.edu/home/viola/. 5

[VW08] Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. *Theory of Computing*, 4:137–168, 2008. 1, 2, 3, 8

A On the seed length and structure of generators for degree-d polynomials

In this section we present two negative results on generators that fool degree-d polynomials which complement our positive results. Before discussing these negative results, let us have a closer look at the seed length of our generator. First, we note that there are "explicit" generators $G : \mathbb{F}^s \to \mathbb{F}^n$ that fool linear polynomials with error ϵ that have seed length

$$s = \log_{|\mathbb{F}|} n + O(\log_{|\mathbb{F}|} 1/\epsilon) + O(1).$$

(We comment below on what we mean by "explicit".) These parameters can be obtained with a probabilistic (i.e., non-explicit) construction, while an "explicit" construction for example follows with some work from the coding-theoretic result in [ABN+92], also using the standard connection between codes and generators for linear polynomials (e.g., [NN93, AGHP92]). We omit the details. Using such a construction, our generator for degree-d polynomials has seed length

$$s = d \cdot \log_{|\mathbb{F}|} n + f(\epsilon, |\mathbb{F}|, d), \tag{12}$$

for some function f. In other words, for fixed $\epsilon, \mathbb{F}$, and d, our seed length is $d \cdot \log_{|\mathbb{F}|} n + O(1)$, as opposed to $O(d \cdot \log_{|\mathbb{F}|} n) + O(1)$ as stated in the introduction. We chose not to present these better parameters in the introduction because these constructions (e.g. [ABN+92]) do not seem to achieve the same explicitness of other constructions (e.g. Lemma 6): It is not obvious how to compute them in time polynomial in $|\mathbb{F}|$ and s. We believe that such strong parameters can be achieved by constructions that are as explicit as the one in Lemma 6, but to our knowledge this has not yet been pointed out.

After having discussed the seed length of our generator, we turn to the discussion of the two results in this section. First, we prove a lower bound on the seed length of generators that fool degree-d polynomials. Informally, the lower bound establishes that the seed length of a generator that fools degree-d polynomials must be at least $d \cdot \log_{|\mathbb{F}|} n - O(1)$. This bound shows that the dependence on n of the seed length of our generator is optimal up to an *additive* term (cf. Equation 12). We derive as a corollary that there is a generator that fools linear polynomials with error tending to 0 (as n grows) such that the sum of $d-1$ independent copies of the generator does *not* fool degree d polynomials with error tending to 0. This result shows that our approach of taking d copies of generators for linear polynomials is *necessary*: $d-1$ copies are not enough. Subsequently to our work, Alon, Ben-Eliezer, and Krivelevich [ABEK08] have derived a more general lower bound on the seed length with a tighter dependence on the error ϵ.

Second, we prove a finer result for the case of quadratic polynomials. We show a non-explicit generator for linear polynomials with exponentially small error that does *not* fool quadratic polynomials – not even up to error 0.9. We point out that this result is incomparable to the previous one, because a linear generator with exponentially small error will have seed length $s \approx n$.

The proofs of these results are fairly straightforward; we present them for completeness.

Claim 30 (Lower bound on the seed length)**.** *Every generator $G : \mathbb{F}^s \to \mathbb{F}^n$ that fools degree-d polynomials with error $|\mathbb{F}|^{-d-3}$ must have seed length $s \geq \log_{|\mathbb{F}|}\left(\binom{n}{1} + \ldots + \binom{n}{d}\right)$. In particular, $s \geq d \cdot \log_{|\mathbb{F}|} n - d \cdot \log_{|\mathbb{F}|} d$.*

Proof. The lemma follows by a dimension counting argument. Let us consider the system of equations

$$p(w) = 0$$

where p is an indeterminate degree-d polynomial and $w \in \mathbb{F}^n$ ranges over all $|\mathbb{F}|^s$ outputs of G. This is a linear system of equations if we think of the coefficients of p as indeterminates.

The number of equations in the system equals the number of all possible outputs of G, which is $|\mathbb{F}|^s$. On the other hand, the number of coefficients of p is at least $\binom{n}{1} + \ldots + \binom{n}{d}$. So if $|\mathbb{F}|^s < \binom{n}{1} + \ldots + \binom{n}{d}$ the system has a non-zero solution, which means that there is a non-zero polynomial p such that $p(G(x)) = 0$ for every $x \in \mathbb{F}^s$.

On the other hand, since p is non-zero, by the Schwartz-Zippel lemma (for small fields) we have $\Pr_X[p(X) = 0] \leq 1 - |\mathbb{F}|^{-d}$. This means that the statistical distance between $p(X)$, $x \in \mathbb{F}^n$, and $p(G(X))$, $x \in \mathbb{F}^s$ is at least $|\mathbb{F}|^{-d}$ and contradicts our assumption that G $|\mathbb{F}|^{-d-3}$-fools degree-d polynomials (using Claim 33).

The last bound in the statement follows from the standard inequality $\binom{n}{d} \geq (n/d)^d$. □

As a corollary, we obtain that there is a generator that fools linear polynomials with error tending to 0 (as n grows) such that the sum of $d-1$ independent copies of the generator does *not* fool degree d polynomials with error tending to 0.

Corollary 31. *There exists a linear generator that $o(1)$-fools linear polynomials over $\mathbb{F}^n$ but such that the sum of $d-1$ independent copies of the generator does not $o(1)$-fool degree d polynomials over $\mathbb{F}^n$.*[7]

Proof sketch. Consider a random generator, or the one that can be derived from [ABN+92]. Such a generator ϵ-fools linear polynomials with seed $s = \log_{|\mathbb{F}|} n + O(\log_{|\mathbb{F}|} 1/\epsilon) + O(1)$. By a suitable choice of $\epsilon = o(1)$, $d-1$ copies of this generator will have seed length shorter than the bound in Claim 30, and the result follows. □

We now present our second result: There is a generator that fools linear polynomials with exponentially small error but that does not fool quadratic polynomials.

Claim 32 (Generators for linear polynomials do not fool quadratic polynomials)**.** *For every n there exists a distribution W that $|\mathbb{F}|^{-\Omega(n)}$-fools linear polynomials over $\mathbb{F}^n$ but that does not (0.9)-fool quadratic polynomials.*

[7] By saying that G "$o(1)$-fools" we mean that there is an $\epsilon = \epsilon(n) = o(1)$ such that G ϵ-fools.

Proof. To simplify notation we will assume n is even; the calculation for odd n is practically identical. Consider the quadratic polynomial

$$p(x_1, \ldots, x_n) := x_1 x_2 + \cdots + x_{n-1} x_n,$$

often called *inner product*. It can be verified that, for uniformly random X, $|\mathrm{E}_X\, e\left[p(X)\right]| \leq |\mathbb{F}|^{-n/4}$.

Define W to be the uniform distribution over the set of all $w \in \mathbb{F}^n$ such that $p(w) = 0$. Clearly, W does not 0.9-fool quadratic polynomials. Thus to complete the proof we only need to show that W fools linear polynomials.

Let l be any nonzero linear function. Then for a random $X \sim \mathbb{F}^n$ and every $i \in \mathbb{F}$ (i denotes a field element, not the complex square root of -1)

$$\begin{aligned}\mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right] &= \sum_{t \in \mathbb{F}} \mathrm{E}_X\, e\left[l(X) + i \cdot t \mid p(x) = t\right] \Pr[p(X) = t] \\ &= \sum_{t \in \mathbb{F}} e(i \cdot t) \cdot \mathrm{E}_X\, e\left[l(X) \mid p(x) = t\right] \Pr[p(X) = t].\end{aligned}$$

We can see the above set of equations (one for every i) as stating that $I = F \cdot T$, where I is the vector of $\mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right]$ (over i), F is the Fourier matrix ($F_{it} = e(i \cdot t)$), and T is the vector of $\mathrm{E}_X\, e\left[l(X) \mid p(x) = t\right] \Pr[p(X) = t]$ (over t). By the inverse Fourier formula, for every $t \in \mathbb{F}$,

$$\mathrm{E}_X\, e\left[l(X) \mid p(x) = t\right] \Pr[p(X) = t] = |\mathbb{F}| \cdot \sum_{i \in \mathbb{F}} e(-i \cdot t) \cdot \mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right].$$

Specializing to $t = 0$, in which case $e(-i \cdot t) = 1$, we have

$$\mathrm{E}_X\, e\left[l(X) \mid p(x) = 0\right] = |\mathbb{F}| \cdot \frac{\sum_{i \in \mathbb{F}} \mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right]}{\Pr[p(X) = 0]}. \tag{13}$$

We now bound (the norm of) $\mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right]$ and $\Pr[p(X) = 0]$. When $i = 0$, $\mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right] = 0$ and when $i \neq 0$, using Equation (4) on Page 8 and Item (3) in Fact 16, we have

$$|\mathrm{E}_X\, e\left[l(X) + i \cdot p(X)\right]| \leq U_2\left(i \cdot p + l\right)^{\Omega(1)} = U_2\left(p\right)^{\Omega(1)} = |\mathbb{F}|^{-\Omega(n)}.$$

It is also not hard to see that $\Pr[p(X) = 0]$ is exponentially close to $1/|\mathbb{F}|$: when $x_1 = x_3 = x_5 = \ldots = 0$ then $p(X) = 0$, otherwise $p(X)$ is uniformly distributed in $\mathbb{F}$. In particular $\Pr[p(X) = 0] \geq 1/(2 \cdot |\mathbb{F}|)$.

Substituting in equation (13) we obtain

$$\left|\mathrm{E}_X\, e\left[l(X) \mid p(X) = 0\right]\right| \leq |\mathbb{F}| \cdot \frac{|\mathbb{F}| \cdot |\mathbb{F}|^{-\Omega(n)}}{(2 \cdot |\mathbb{F}|)^{-1}} \leq |\mathbb{F}|^{-\Omega(n)}.$$

□

B Statistical distance and character distance

In this section we point that our results also apply to the definition of pseudorandomness in terms of statistical distance, as opposed to our algebraic Definition 4.

Let $\mathbb{F}$ be an arbitrary (not necessarily prime) field. If X and X' are random variables taking values in $\mathbb{F}$, their *statistical distance* is the quantity

$$\begin{aligned}\mathrm{SD}(X,X') &= \frac{1}{2}\sum_{x\in\mathbb{F}}\left|\Pr_X[X=x]-\Pr_{X'}[X'=x]\right| \\ &= \max_{S\subseteq\mathbb{F}}\left(\Pr_X[X\in S]-\Pr_{X'}[X'\in S]\right).\end{aligned}$$

In contrast, our notion of pseudorandomness (Definition 4) is given in terms of the following quantity, which we think of as (maximum) *character distance*.

$$\mathrm{CD}(X,X') = \max_{a\in\mathbb{F}}\left|\mathrm{E}_X[e_a(X)]-\mathrm{E}_{X'}[e_a(X')]\right|. \tag{14}$$

The statistical distance and character distance are related as follows.

Claim 33. *For every random variables X and X' taking values in $\mathbb{F}$ we have*

$$\mathrm{CD}(X,X') \le 2\cdot\mathrm{SD}(X,X') \le \sqrt{|\mathbb{F}|-1}\cdot\mathrm{CD}(X,X').$$

Theorems 1 and 2 show that when the field $\mathbb{F}$ is prime than for all polynomials p of degree d, the quantity $\mathrm{E}[e_1(p(X))]-\mathrm{E}_W[e_1(p(W))]$ is small, where X is uniform and W is pseudorandom. Using the fact that $e_a(x)=e(ax)$ for every $a,x\in\mathbb{F}$, it follows that $p(X)$ and $p(W)$ are close in character distance, so by Claim 33 they are also close in statistical distance.

We give a simple extension of this argument which works for general (not necessarily prime) fields $\mathbb{F}$ in Appendix C.

Proof of Claim 33. Let $p(x)=\Pr_X[X=x]$ and $p'(x)=\Pr_{X'}[X'=x]$.

For the first inequality, we have

$$\left|\mathrm{E}_X\, e_a(X)-\mathrm{E}_{X'}\, e_a(X')\right| = \left|\sum_{x\in\mathbb{F}}(p(x)-p'(x))\cdot e_a(x)\right| \le \sum_{x\in\mathbb{F}}|p(x)-p'(x)| = 2\cdot\mathrm{SD}(X,X').$$

For the other one, let $\Delta=\mathrm{CD}(X,X')$. We think of p and p' as functions from $\mathbb{F}$ to $\mathbb{C}$. Then equation (14) says that for all $a\in\mathbb{F}$,

$$|\hat{p}_a-\hat{p}'_a| = \frac{1}{|\mathbb{F}|}\cdot\left|\mathrm{E}_X[e_{-a}(X)]-\mathrm{E}_{X'}[e_{-a}(X')]\right| \le \frac{\Delta}{|\mathbb{F}|}.$$

Applying Parseval's identity (see Section 4.2) to the function $p-p'$, we have that

$$\frac{1}{|\mathbb{F}|}\sum_{x\in\mathbb{F}}(p(x)-p'(x))^2 = \sum_{a\in\mathbb{F}}(\hat{p}_a-\hat{p}'_a)^2 \le (|\mathbb{F}|-1)\cdot\frac{\Delta^2}{|\mathbb{F}|^2},$$

where in the last inequality we use the fact that $\hat{p}_0 - \hat{p}'_0 = 1 - 1 = 0$.

It now follows by the Cauchy-Schwarz inequality that

$$\left(2 \cdot \frac{\mathrm{SD}(X, X')}{|\mathbb{F}|}\right)^2 = \left(\frac{1}{|\mathbb{F}|}\sum_{x\in\mathbb{F}}|p(x) - p'(x)|\right)^2 \leq \frac{1}{|\mathbb{F}|}\sum_{x\in\mathbb{F}}(p(x) - p'(x))^2 \leq \frac{|\mathbb{F}| - 1}{|\mathbb{F}|^2} \cdot \Delta^2. \quad \square$$

C Extension to non-prime fields

We now argue that our main results can be extended to the case of non-prime fields. Let q be a prime and consider the field $\mathbb{K} = \mathbb{F}_{q^t}$, where t is an arbitrary integer. Let γ be a root of some irreducible polynomial of degree t over $\mathbb{F} = \mathbb{F}_q$. Then every element a of $\mathbb{F}_{q^t}$ can be written as $a_0 + a_1\gamma + \cdots + a_{t-1}\gamma^{t-1}$. We represent a by the t-tuple $(a_0, \ldots, a_{t-1}) \in \mathbb{F}^t$. In this representation, an input $x \in \mathbb{K}^n$ can be viewed as element of the space $\mathbb{F}^{nt}$.

We prove that the problem of fooling polynomials over $\mathbb{K}$ reduces to the problem of fooling polynomials of the same degree over $\mathbb{F}$. Specifically:

Claim 34. *Let X, X' be any pair of distributions over $\mathbb{F}^{nt}$ such that*

$$|\mathrm{E}_X\, e\left[r(X)\right] - \mathrm{E}_{X'}\, e\left[r(X')\right]| < \epsilon$$

for all polynomials r of degree d over $\mathbb{F}^{nt}$. Let $p : \mathbb{K}^n \to \mathbb{K}$ be any degree d polynomial. Then for every character e_a of $\mathbb{K}$,

$$\left|\mathrm{E}_X[e_a(p(X))] - \mathrm{E}_{X'}[e_a(p(X'))]\right| < \epsilon.$$

By Claim 33, it then follows that $p(X)$ and $p(X')$ are $\epsilon \cdot \sqrt{|\mathbb{F}| - 1}$-close in statistical distance.

Proof. We can write $p(x) = p_0(x) + p_1(x) \cdot \gamma + \cdots + p_{t-1}(x) \cdot \gamma^{t-1}$, where we view each p_i as a function from $\mathbb{F}^{nt} \to \mathbb{F}$. If p is of degree d, then each p_i is also of degree at most d, so it follows that the function $\langle a, p(x)\rangle = a_0p_0(x) + \cdots + a_{t-1}p_{t-1}(x)$ is also a polynomial of degree d from $\mathbb{F}^{nt} \to \mathbb{F}$. Setting $r(x) = \langle a, p(x)\rangle$ concludes the proof. $\square$