Electronic Colloquium on Computational Complexity, Report No. 48 (2009)

On the Complexity of Boolean Functions in Different Characteristics

Parikshit Gopalan
Microsoft Research-Silicon Valley
parik@microsoft.com

Shachar Lovett [*]
The Weizmann Institute of Science
shachar.lovett@weizmann.ac.il

Amir Shpilka [†]
The Technion - Israel Institute of Technology
shpilka@cs.technion.ac.il

May 29, 2009

Abstract

Every Boolean function on n variables can be expressed as a unique multivariate polynomial modulo p for every prime p. In this work, we study how the degree of a function in one characteristic affects its complexity in other characteristics. We establish the following general principle: *functions with low degree modulo p must have high complexity in every other characteristic q.* More precisely, we show the following results about Boolean functions $f : \{0,1\}^n \to \{0,1\}$ which depend on all n variables, and distinct primes p, q:

- If f has degree $o(\log n)$ modulo p, then it must have degree $\Omega(n^{1-o(1)})$ modulo q. Thus a Boolean function has degree $o(\log n)$ in only one characteristic. This result is essentially tight as there exist functions that have degree $\log n$ in every characteristic.
- If f has degree $d = o(\log n)$ modulo p, it cannot be computed correctly on more than $1 - p^{-O(d)}$ fraction of the hypercube by polynomials of degree $n^{\frac{1}{2}-\epsilon}$ modulo q.

As a corollary of the above results it follows that if f has degree $o(\log n)$ modulo p, then it requires super-polynomial size $AC_0[q]$ circuits. This gives a lower bound for a broad and natural class of functions.

[*] Research supported by the Israel Science Foundation (grant 1300/05).
[†] Research supported by the Israel Science Foundation (grant 439/06).

ISSN 1433-8092

1 Introduction

Representations of Boolean functions as polynomials in various characteristics have been studied intensively in Computer science [NS92, Pat92, Bei93, BBR94]. This algebraic view of Boolean functions has found numerous applications to diverse areas including circuit lower bounds [Raz87, Smo87, BRS91, ABFR94], computational learning [KM93, LMN93, KS01, MOS03] and explicit combinatorial constructions [Gro00, Gro02, Gop06b, Efr09]. As a purely algebraic model of computation, polynomial representations lead to some natural complexity measures such as exact degree, approximation degree and sparsity needed to represent a function. In this work, we are primarily concerned with the polynomial degree of a function, defined as follows:

Definition 1.1. *For a Boolean function $f : \{0,1\}^n \to \{0,1\}$, the degree of f in characteristic k, denoted $\deg_k(f)$, is the degree of the unique multilinear polynomial $P(X_1,\ldots,X_n) \in R[X_1,\ldots,X_n]$ such that $P(x) = f(x)$ for every $x \in \{0,1\}^n$, where $R = \mathbb{Z}/k\mathbb{Z}$.*

We say that the polynomial P represents f over R. The existence and uniqueness of such a representing polynomial follows from the Möbius inversion formula (see Section 2). Of particular importance in complexity theory are the cases $k = 0$ ($R = \mathbb{Z}$) and $k = p$ ($R = \mathbb{Z}_p$) for some prime p; these will also be our primary focus, though we will also consider the case of composite m. We denote $\deg_0(f)$ simply by $\deg(f)$; it also equals the degree of the Fourier polynomial for the function $(-1)^{f(x)}$. Let us note a basic relation between these various degrees, namely that for every f and p, we have

$$\deg_p(f) \leq \deg(f).$$

This is because the polynomial representing f over $\mathbb{Z}_p$ can be obtained from the representation over $\mathbb{Z}$ by taking each coefficient modulo p. The gap between these quantities can be arbitrarily large; consider the function $\mathsf{Parity}(x) = \sum_i x_i \bmod 2$. It is easy to show that $\deg(\mathsf{Parity}) = n$ whereas $\deg_2(\mathsf{Parity}) = 1$. Indeed, it is not hard to show that $\deg_p(\mathsf{Parity}) = n$ for every $p \neq 2$.

In this paper, we show that this is an instance of a more general principle:

A function on all n variables which has low degree in characteristic p is bound to have high degree in every other prime characteristic $q \neq p$.

Moreover, we prove that any function f where $\deg_p(f) = o(\log n)$ is hard to approximate by low-degree polynomials modulo q, and hence requires large $AC_0[q]$ circuits.

1.1 Our Results

When we refer to Boolean functions on n variables, we only consider functions where all n variables are influential. This rules out trivial counterexamples like k-juntas that have low degree in all characteristics. The following is our main theorem:

Theorem 1.2. *(Main) Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which depends on all n variables. Let $p \neq q$ be distinct primes. Then*

$$\deg_q(f) \geq \frac{n}{\lceil \log_2 p \rceil \deg_p(f) p^{2\deg_p(f)}}.$$

This gives a lower bound of $\Omega(n^{1-o(1)})$ on $\deg_q(f)$ as long as $\deg_p(f) = o(\log n)$. This bound is close to the best possible, as there exist functions on all n variables (such as the addressing function [NS92]) where $\deg(f) \leq \log n$, and hence $\deg_p(f) \leq \log n$ for all characteristics p. Thus one cannot get nontrivial lower bounds on $\deg_q(f)$ once $\deg_p(f)$ exceeds $\log n$.

Nisan and Szegedy show that any function on n variables must have degree at least $\deg(f) \geq \log n - O(\log \log n)$ [NS92]. An interesting consequence of Theorem 1.2 is the following analogue of the Nisan-Szegedy bound for non-prime power moduli.

Corollary 1.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which depends on all n variables. Suppose $p < q$ are distinct primes that divide m. We have that*

$$\deg_m(f) \geq \frac{1}{2}\log_p n - \log_p \log_p n.$$

This corollary is interesting as it illuminates a sharp difference between degrees over composite numbers and over primes. A simple way to construct functions which have degree $O(1)$ over $\mathbb{F}_p$ is to take any constant degree polynomial $P(x_1, \ldots, x_n) \in \mathbb{F}_p[x_1, \ldots, x_n]$ and raise it to the power $p-1$. This construction fails for composite m since there is no analogue of Fermat's little theorem. Corollary 1.3 shows that indeed any polynomial modulo m computing a Boolean function requires degree $\Omega(\log n)$, as it does over the reals.

While Theorem 1.2 also implies lower bounds for $\deg(f)$, one can show a stronger bound by a simple modification of the Nisan-Szegedy proof:

$$\deg(f) \geq \frac{n}{2^{\deg_p(f)}}.$$

The results above show a very basic relation between the degrees of Boolean functions over different characteristic. A natural question to ask is what happens if we relax the requirement and only consider polynomials over $\mathbb{F}_q$ that approximate a low degree polynomial over $\mathbb{F}_p$. However, similarly to the case of degree 1 polynomials that was studied in [Smo87], we prove that low degree polynomials modulo p are hard to even approximate by polynomials in other characteristics.

Theorem 1.4. *Given $f : \{0,1\}^n \to \{0,1\}$ such that $\deg_p(f) = d$, for any $q \neq p$ and $Q(x_1, \ldots, x_n) \in \mathbb{Z}_q[x_1, \ldots, x_n]$ with $\deg(Q) = o(\sqrt{n})$,*

$$\Pr_{x \in \{0,1\}^n}[f(x) = Q(x)] \leq 1 - \epsilon p^{-d},$$

where ϵ depends only on q.

We note that both the error bound of $1-p^{-O(d)}$ and the degree bound of $o(\sqrt{n})$ are close to optimal; there are polynomials of degree d over $\mathbb{Z}_p$ that are 0 with probability $1-2^{-d}$, hence they have trivial approximations over $\mathbb{Z}_q$. Secondly, the Mod_p function (and indeed every symmetric function) can be $1-\epsilon$ approximated by polynomials of degree $c(\epsilon)\sqrt{n}$ over $\mathbb{Z}_q$ [BGL06], despite being hard to approximate for polynomials of lower degree.

As a corollary of Theorem 1.4 we get that if a Boolean function has low degree modulo p, then the function requires large $AC_0[q]$ circuits for any prime $q \neq p$. Several of the known lower bounds for $AC_0[q]$ are functions like Parity and the Mod_{p^k} function where $p \neq q$ that are easily seen to be low-degree polynomials in some characteristic. Our result generalizes this to give a very general class of hard functions for $AC_0[q]$, namely all functions that have degree $o(\log n)$ modulo $p \neq q$.

Theorem 1.5. *Let p, q be distinct primes. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function on n variables with $\deg_p(f) = o(\log_p n)$. Then any $AC_0[q]$ circuit of depth t computing f requires size at least $exp(n^{(1-o(1))/2t})$.*

It is not hard to see that most known lower bounds for $AC_0[q]$ circuits follow from the theorem above. For example, the lower bound for Mod_{p^k} of [Smo87] follows from the observation that $\deg_p(\mathsf{Mod}_{p^k}) \leq p^k$ (see e.g. [BGL06]). Additionally, it gives several new lower bounds, for instance it shows that every quadratic form on n variables over $\mathbb{F}_2$ requires large $AC_0[q]$ circuits, for $q \neq 2$. Though we note that Theorem 1.5 does not imply Razborov's lower bound for $\mathsf{Majority}$.

Summarizing, Theorems 1.2 and 1.4 show that for a Boolean function, having low degree mod p, or even being close to a low degree polynomial mod p, is a "singular" event, in the sense it can only occur for at most one characteristic p.

1.2 Polynomial representations in computer science.

The study of polynomial representations of Boolean functions dates at least as far back as the 1960's, when they arose in various contexts including switching theory [Mur71], voting theory [Cho61] and machine learning [MP68]. Representations of Boolean functions over finite fields, especially over $\mathbb{F}_2$ were studied by coding theorists in the context of Reed-Muller codes, see [MS77, Chapters 13-14] and the references therein. The codewords of the code $\mathsf{RM}(d,n)$ are all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ where $\deg_2(f) \leq d$, while received words are arbitrary functions f.

Polynomial representations have proved especially useful in circuit complexity [Bei93] where a natural lower bound technique is to relate concrete complexity measures (such as circuit-size) which we wish to bound, to purely algebraic complexity measures. Examples of this paradigm include the Razborov-Smolensky lower bounds for $AC_0[p]$ [Raz87, Smo87], which relates the circuit size to the polynomial degree needed to approximate f over $\mathbb{Z}_p$, and the work of Beigel *et al.* [BRS91] and Aspnes *et al.* [ABFR94] which relate AC_0 circuit size with approximations by real polynomials.

Polynomial representations are among the most powerful tools in computational learning. The best learning algorithms for many basic concept classes, including but not limited to decision trees [KM93], DNF formulae [KS01], AC_0 circuits [LMN93, JKS02], juntas [MOS03] and halfspaces [KOS02, KKMS05] all proceed by showing that the concept class to be learned has some *nice* polynomial representation. In particular, the algorithm for learning juntas of [MOS03] exploits a connection between $\deg_2(f)$ and the sparsity of its Fourier polynomial.

Finally, polynomial representations of Boolean functions have found applications to constructing combinatorial objects such as set systems [Gro00, Gro02], Ramsey graphs [Gro00, Gop06b] and locally decodable codes [Efr09]. These results require low-degree *weak* representations of simple Boolean functions like the Or function but modulo composites.

Definition 1.6. *The polynomial $P(x_1, \ldots, X_n) \in \mathbb{Z}_m[X_1, \ldots, X_n]$ weakly represents $f : \{0,1\}^n \to \{0,1\}$ over $\mathbb{Z}_m$ if $f(x) \neq f(y) \Rightarrow P(x) \neq P(y)$. ($P(x)$ may take values in $\mathbb{Z}_m$)*

Such representations have been well studied in complexity theory (see [BBR94, BGL06] and the references therein), but embarrassingly simple questions like the degree required to represent the Or function mod 6 remain open, there is a gap of $O(\sqrt{n})$ [BBR94] versus $\Omega(\log n)$ [TB98] between upper and lower bounds. Better upper bounds would lead to improved constructions of all the above combinatorial objects. In [Gop06b], Gopalan proposes viewing this as a question about the degree of two related functions in distinct characteristics:

Problem 1.7. *[Gop06b] If two functions $f, g : \{0,1\}^n \to \{0,1\}$ satisfy $f(x) \vee g(x) = \mathsf{Or}(x)$, how small can $\max(\deg_2(f), \deg_3(g))$ be?*

Questions like this emphasize the importance of the natural and basic question of understanding the behavior of $\deg_p$ for various characteristics p.

1.3 Techniques.

Our proofs are conceptually very simple, we reduce the degree d case to the linear case and then appeal to known lower bounds. This reduction is carried out via a degree reduction lemma (Lemma 3.1) that shows that for any degree d polynomial $P(x)$ over $\mathbb{Z}_p$ on n variables, there exists a constant t and a linear combination of the form

$$P'(x) = \sum_{i \leq t} \lambda_i P(x + a_i) \quad \lambda_i \in \mathbb{Z}_p, \; a_i \in \mathbb{Z}_p^n$$

so that by fixing some variables in P' to constants, we get a linear polynomial in many variables. This lemma is proved using discrete derivatives, a notion that has proved very useful lately in complexity theory [BV07, Lov08, Vio08].

With this lemma in hand, one would like to proceed as follows: suppose $P(x)$ and $Q(x)$ represent the same function f over $\mathbb{Z}_p$ and $\mathbb{Z}_q$, and that $P(x)$ has low degree (say a constant). We would like to claim that the degree of $P'(x)$ over $\mathbb{Z}_q$ is a small multiple of $\deg(Q)$, which would then imply that $\deg(Q)$ must be large, since the Mod_p function has high degree in characteristic q. Implementing this scheme runs into many obstacles: P' is a function that maps $\mathbb{Z}_p^n \to \mathbb{Z}_p$, further the values a_i are from $\mathbb{Z}_p^n$, thus while $P(x) = Q(x)$ for $x \in \{0,1\}^n$, it is unclear how $Q(x)$ can help us evaluate $P(x+a_i)$. Most of the technical work in this paper goes towards circumventing this obstacle, and showing that one can mimic differentiation modulo p in characteristic q without a large blowup in the degree. Note that in the case when $p=2$, these complications do not arise (since $\{0,1\}^n \subset \mathbb{Z}_q^n$), making the proofs much simpler. So we present the case of characteristic 2 separately in Section 4, and the general case in Section 5.

2 Preliminaries

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. We will only consider Boolean functions that depend on all n variables, meaning that they cannot be written as $f(x_1,\ldots,x_n) = g(x_{i_1},\ldots,x_{i_k})$ for some $k < n$. We start by establishing the correspondence between functions and polynomials. We state the correspondence in the general setting of any commutative ring R containing $\{0,1\}$, but we will only be interested in the cases where R is either $\mathbb{Z}$, $\mathbb{Z}_m$ for some integer m or a finite field $\mathbb{F}_q$. We say that a polynomial $P(x_1,\ldots,x_n) \in R[x_1,\ldots,x_n]$ computes the function f if $P(x) = f(x)$ for all $x \in \{0,1\}^n$. While there could be many polynomials that satisfy this condition, if we insist that the polynomial be multilinear (every variable occurs with degree at most 1), then the polynomial is unique. This can be seen via the Möbius inversion formula, which gives a unique multilinear polynomial $P(x_1,\ldots,x_n) \in R[x_1,\ldots,x_n]$ satisfying $P(x) = f(x)$ for every function $f : \{0,1\}^n \to R$:

$$\begin{aligned} P(x) &= \sum_{S\subseteq[n]} c_S \prod_{i\in S} x_i \\ \text{where} \quad c_S &= \sum_{x\le x(S)} (-1)^{|S|-wt(x)} f(x) \end{aligned}$$

where $x(S)$ denotes the indicator vector of the set S, $x \le x(S)$ denotes that $x_i \le x(S)_i$ for every coordinate i and $wt(x)$ denotes the Hamming weight of the vector x. If f is Boolean, the Möbius inversion shows that the representing polynomial depends only on the characteristic of R.

We state some basic facts about $\deg_k(f)$, proofs of which can be found in [Gop06a]. The multilinear polynomial computing f over $\mathbb{Z}_m$ can be obtained by reducing each coefficient of the polynomial computing f over $\mathbb{Z}$ modulo m, which gives the following:

Fact 2.1. *For any $f : \{0,1\}^n \to \{0,1\}$, we have $\deg_m(f) \le \deg(f)$ for all m. Similarly if $m_1|m$, then $\deg_{m_1}(f) \le \deg_m(f)$.*

A consequence of this inequality is that $\deg_m(f) \le \deg_{m^k}(f)$. The following folklore lemma shows that they are always within a factor $2k$ of each other.

Fact 2.2. *For any $f : \{0,1\}^n \to \{0,1\}$, and integers m, k:*

$$\deg_m(f) \le \deg_{m^k}(f) \le (2k-1)\deg_m(f).$$

If $m = m_1 m_2$ where $(m_1, m_2) = 1$, then the multilinear polynomial $P(x) \in \mathbb{Z}_m[x]$ is obtained by combining the coefficients of $P_1(x) \in \mathbb{Z}_{m_1}[x]$ and $P_2(x) = \mathbb{Z}_{m_2}[x]$ by the Chinese Remainder Theorem. Hence

Fact 2.3. *Let $m = m_1 m_2$ where $(m_1, m_2) = 1$. Then*

$$\deg_m(f) = \max(\deg_{m_1}(f), \deg_{m_2}(f)).$$

Thus if we know $\deg_p(f)$ for all primes p that divide m, we can use Facts 2.2 and 2.3 to estimate $\deg_m(f)$ up to a constant factor which is independent of n but depends on m.

We define the function $\mathsf{Mod}_m(x)$ to be 1 whenever $\sum_i x_i$ is divisible by m. The degree of such functions in any characteristic can be computed using the following observation:

Fact 2.4. *For any integer k, and primes $p \neq q$, we have*

$$\deg_p(\mathsf{Mod}_{p^k}) = p^k, \quad \deg_q(\mathsf{Mod}_{p^k}) = \Omega(n).$$

Finally, we use two lemmas from the work of Razborov and Smolensky showing that if a Boolean function f can be computed by a small $AC_0[p]$ circuit, then f can be well approximated by low degree polynomials over $\mathbb{F}_p$. The first is their low-degree approximation lemma for $AC_0[p]$ circuits.

Lemma 2.1 (Razborov-Smolensky [Raz87, Smo87])**.** *Let f be a Boolean function on n variables that is computed by an $AC_0[p]$ circuit of size s and depth t. For every $\delta > 0$, there exists a polynomial $P \in \mathbb{F}_p[x_1, \ldots, x_n]$ of degree $\deg(P) \leq (cp\log(s/\delta))^t$ such that $P(\{0,1\}^n) \subset \{0,1\}$ and*

$$\Pr_{x \in \{0,1\}^n}[P(x) = f(x)] \geq 1 - \delta$$

for some universal constant c.

The second shows that the Mod_p function does not have such approximations over $\mathbb{Z}_q$.

Lemma 2.2 (Razborov-Smolensky [Raz87, Smo87])**.** *For any prime $p \neq q$, there exist constants $c, \epsilon > 0$ such that for any polynomial $Q(x)$ over $\mathbb{Z}_q$ of degree at most $c\sqrt{n}$,*

$$\Pr_{x \in \{0,1\}^n}[Q(x) = \mathsf{Mod}_p(x)] < 1 - \epsilon.$$

3 Degree Reduction

A crucial tool in our proof is the following *Degree reduction lemma* which reduces degree d polynomials in n variables to polynomials with many linear terms. For a polynomial P define the set $L(P)$ of all variables x_i which occur as linear terms, but not in any higher degree monomials.

Lemma 3.1 (Degree Reduction Lemma). *Let $P(x)$ be a polynomial of degree d over $\mathbb{Z}_p$ which depends on all n variables, where the individual degree of each variable is at most $p-1$. There exists $t \leq p^{\lceil \frac{d-1}{p-1} \rceil}$, $a_1, \ldots, a_t \in \mathbb{Z}_p^n$, and $\lambda_1, \ldots, \lambda_t \in \mathbb{Z}_p$ such that the polynomial*

$$Q(x) = \sum_{i \leq t} \lambda_i P(x + a_i)$$

satisfies

$$|L(Q)| \geq \frac{n}{dp^{\lceil \frac{d-1}{p-1} \rceil}}.$$

The reminder of this section is dedicated to the proof of Lemma 3.1. We define the *monomial degree* of a variable x_i in a polynomial $P(x)$ to be the maximal degree of a monomial of P containing x_i, and denote it by $\deg_i(P)$. Note that the monomial degree of x_i is different from its individual degree, which is the highest power of x_i that occurs in P. The main tool we use to prove this lemma is the notion of directional derivatives of a polynomial. Given a polynomial P, we define the first derivative along y, denoted $P_{(y,1)}$, as

$$P_{(y,1)}(x) = P(x+y) - P(x).$$

We define the ℓ^{th} derivative along y for $\ell \geq 1$ inductively as

$$P_{(y,\ell)}(x) = P_{(y,\ell-1)}(x+y) - P_{(y,\ell-1)}(x)$$

when $\ell \geq 1$. It is easy to verify that

$$P_{(y,\ell)}(x) = \sum_{0 \leq j \leq \ell} (-1)^{\ell - j} \binom{\ell}{j} P(x + jy).$$

We define multiple derivatives in multiple directions, which we denote by $P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}(x)$. To write closed forms for such derivatives, we define the following quantity for all ℓ, c:

$$\mu(\ell, c) = \sum_{0 \leq j \leq \ell} (-1)^{\ell - j} \binom{\ell}{j} j^c.$$

The following combinatorial identities are well-known; we prove them for completeness:

Fact 3.1. *Let $\ell \leq p-1$. Then*

$$\mu(\ell, c) = 0 \textit{ for } c \in \{0, \ldots, \ell - 1\},$$
$$\mu(\ell, \ell) \not\equiv 0 \mod p$$

Proof. We prove the first identity by induction on c. The case $c = 0$ is elementary. To prove it for $c \geq 1$, we have

$$(X-1)^\ell = \sum_{0 \leq j \leq \ell} (-1)^{\ell - j} \binom{\ell}{j} X^j$$

Differentiating $c \le \ell - 1$ times and then setting $X = 1$ gives

$$\begin{aligned} 0 &= \sum_{0 \le j \le \ell} (-1)^{\ell - j} \binom{\ell}{j} j(j-1) \cdots (j-c+1) \\ &= \mu(\ell, c) + \sum_{1 \le i \le c-1} \lambda(i) \mu(\ell, i) \\ &= \mu(\ell, c) \end{aligned}$$

where we use the induction hypothesis for $i \le c - 1$ to set $\mu(\ell, i) = 0$ for $i \le c - 1$.

To prove $\mu(\ell, \ell) \not\equiv 0 \mod p$, we differentiate ℓ times to get

$$\begin{aligned} \ell! &= \sum_{0 \le j \le \ell} (-1)^{\ell - j} \binom{\ell}{j} j(j-1) \cdots (j-\ell+1) \\ &= \mu(\ell, \ell) + \sum_{1 \le c \le \ell-1} \lambda(c) \mu(\ell, c) \\ &= \mu(\ell, \ell) \end{aligned}$$

Since we assume $\ell \le p - 1$, we have $\mu(\ell, \ell) = \ell! \not\equiv 0 \bmod p$. □

We abbreviate the monomial $\prod_{i=1}^{n} x_i^{d_i}$ by x^d where $d = (d_1, \cdots, d_n)$ is the degree vector. We use $|d| = \sum_i d_i$ to denote its total degree. Given vectors d, e we use the notation $\binom{d}{e} = \prod_i \binom{d_i}{e_i}$. We have

$$\begin{aligned} x^d_{(y,\ell)} &= \sum_{j=0}^{\ell} (-1)^{\ell-j} \binom{\ell}{j} (x + jy)^d \\ &= \sum_{j=0}^{\ell} (-1)^{\ell-j} \binom{\ell}{j} \sum_{e \le d} \binom{d}{e} x^{d-e} (jy)^e \\ &= \sum_{e \le d} \binom{d}{e} x^{d-e} y^e \sum_{j=0}^{\ell} (-1)^{\ell-j} \binom{\ell}{j} j^{|e|} \\ &= \sum_{e \le d} \binom{d}{e} x^{d-e} y^e \mu(\ell, |e|) \\ &= \sum_{\substack{e \le d \\ |e| \ge \ell}} \binom{d}{e} x^{d-e} y^e \mu(\ell, |e|) \end{aligned}$$

where we use $\mu(\ell, |e|) = 0$ for $|e| \le \ell - 1$. Thus, differentiating ℓ times along y reduces the degree in x by at least ℓ, as one would expect.

By repeating this calculation, we can compute an expression for derivatives in multiple directions. Given vectors $d, e^{(1)}, \ldots, e^{(k)}$ we use the notation $\binom{d}{e^{(1)}, \ldots, e^{(k)}}$ for the product of multinomials $\prod_{l\in[n]}\binom{d_l}{e_l^{(1)},\dots,e_l^{(k)}}$. We have

$$\begin{aligned}
&x^d_{(y^{(1)},\ell^{(1)}),\dots,(y^{(k)},\ell^{(k)})} = \\
&= \sum_{e^{(1)}+\cdots+e^{(k)}\leq d} \binom{d}{e^{(1)},\dots,e^{(k)}} x^{d-(e^{(1)}+\cdots+e^{(k)})}\cdot \\
&\qquad \cdot \prod_{j=1}^{k} \mu(\ell^{(j)},|e^{(j)}|)(y^{(j)})^{e^{(j)}} \\
&= \sum_{\substack{e^{(1)}+\cdots e^{(k)}\leq d\\ |e^{(1)}|\geq \ell^{(1)},\dots,|e^{(k)}|\geq \ell^{(k)}}} \binom{d}{e^{(1)},\dots,e^{(k)}} x^{d-(e^{(1)}+\cdots+e^{(k)})}\cdot \\
&\qquad \cdot \prod_{j=1}^{k} \mu(\ell^{(j)},|e^{(j)}|)(y^{(j)})^{e^{(j)}}
\end{aligned}$$

By linearity, we can compute the derivative of any polynomial $P(x) = \sum_d c_d x^d$.

$$\begin{aligned}
&P_{(y^{(1)},\ell^{(1)}),\dots,(y^{(k)},\ell^{(k)})}(x) = \sum_d c_d x^d_{(y^{(1)},\ell^{(1)}),\dots,(y^{(k)},\ell^{(k)})} \\
&= \sum_d c_d \sum_{\substack{e^{(1)}+\cdots e^{(k)}\leq d\\ |e^{(1)}|\geq \ell^{(1)},\dots,|e^{(k)}|\geq \ell^{(k)}}} \binom{d}{e^{(1)},\dots,e^{(k)}} x^{d-(\sum_j e^{(j)})}\cdot \\
&\qquad \cdot \prod_{j=1}^{k} \mu(\ell^{(j)},|e^{(j)}|)(y^{(j)})^{e^{(j)}} \\
&= \sum_f x^f \Big(\sum_{|e^{(1)}|\geq \ell^{(1)},\dots,|e^{(k)}|\geq \ell^{(k)}} c_{f+\sum_j e^{(j)}} \binom{f+\sum_j e^{(j)}}{e^{(1)},\dots,e^{(k)}}\cdot \\
&\qquad \cdot \prod_{j=1}^{k} \mu(\ell^{(j)},|e^{(j)}|)(y^{(j)})^{e^{(j)}} \Big) \qquad (1)
\end{aligned}$$

where in the last line we use the change of variable $f = d - \sum_j e^{(j)}$. Recall that we define $\deg_i(P)$ to be the largest degree monomial containing the variable x_i. It follows that the monomial degree of x_i drops by at least $\sum_j \ell^{(j)}$:

$$\deg_i(P_{(y^{(1)},\ell^{(1)}),\dots,(y^{(k)},\ell^{(k)})}) \leq \deg_i(P) - \sum_j \ell^{(j)}.$$

Lemma 3.2. *Let*

$$\begin{gathered}
\deg_i(P) = (k-1)(p-1)+\ell+1 \quad \textit{where} \quad \ell+1\leq p-1, \\
\ell^{(1)} = \cdots = \ell^{(k-1)} = p-1 \text{ and } \ell^{(k)} = \ell.
\end{gathered}$$

Then the coefficient of x_i in $P_{(y^{(1)},\ell^{(1)}),\dots,(y^{(k)},\ell^{(k)})}(x)$ is a non-zero polynomial in $y^{(1)},\dots,y^{(k)}$.

Proof. Observe that $\sum_j \ell^{(j)} = \deg_i(P) - 1$, so

$$\deg_i(P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}) \leq \deg_i(P) - \sum_j \ell^{(j)} = 1.$$

Our goal is to show that it is in fact 1. Take the vector f where $f_i = 1$ and $f_j = 0$ for all $j \neq i$. By Equation 1, the coefficient of x^f is given by

$$c'_f = \sum_{e^{(1)},\ldots,e^{(k)}} c_{f+\sum_j e^{(j)}} \binom{f + \sum_j e^{(j)}}{e^{(1)},\ldots,e^{(k)}}.$$
$$\prod_{j=1}^{k} \mu(\ell^{(j)}, |e^{(j)}|)(y^{(j)})^{e^{(j)}}$$

Our goal is now to find $e^{(1)},\ldots,e^{(k)}$ so that the following conditions hold:

$$c_{f+\sum_j e^{(j)}} \neq 0, \quad \binom{f + \sum_j e^{(j)}}{e^{(1)},\ldots,e^{(k)}} \neq 0 \tag{2}$$
$$|e^{(1)}| = \cdots = |e^{(k-1)}| = p-1, |e^{(k)}| = \ell \tag{3}$$

Equation (3) ensures that $\mu(\ell_j, |e^{(j)}|) \neq 0$. So each solution $(e^{(1)},\cdots,e^{(k)})$ will contribute a non-zero multiple of the monomial $\prod_{j=1}^k (y^{(j)})^{e^{(j)}}$ to c'_f, and distinct solutions will contribute distinct monomials. Thus the claim follows if we show that there is at least one solution.

Fix a monomial x^d, where $|d| = \deg_i(P)$ and $c_d \neq 0$, which contains the variable x_i. Now $|d - f| = (k-1)(p-1) + \ell$. It is easy to define $e^{(1)},\ldots,e^{(k)}$ so that

$$|e^{(1)}| = \cdots = |e^{(k-1)}| = p-1, |e^{(k)}| = \ell$$
$$\sum_j e_l^{(j)} + f_l = d_l \quad \forall\ l \in [n]$$

Note that

$$\binom{f + \sum_j e^{(j)}}{e^{(1)},\ldots,e^{(k)}} = \prod_{l\in[n]} \binom{f_l + \sum_j e_l^{(j)}}{e_l^{(1)},\ldots,e_l^{(k)}}.$$

Since

$$\sum_j e_l^{(j)} \leq f_l + \sum_j e_l^{(j)} \leq d_l \leq p-1$$

each binomial coefficient in the product is non-zero mod p. This gives a solution satisfying both Equations 2 and 3. □

Let $\delta_p(d)$ denote the minimum probability that a degree d polynomial over $\mathbb{Z}_p$ is non-zero. It is well-known (see e.g. [MS77]) that if $d = a(p-1) + b$ where $a \geq 0$ and $b \leq p-1$, then

$$\delta_p(d) = \frac{1}{p^a}\left(1 - \frac{b}{p}\right) \geq p^{-\lceil \frac{d}{p-1} \rceil}$$

Lemma 3.3. *Let $P(x) \in \mathbb{Z}_p[x]$ be a degree d polynomial that depends on all n variables. Then there exists $k \leq \lceil \frac{d-1}{p-1} \rceil$, directions $y^{(1)}, \ldots, y^{(k)} \in \mathbb{Z}_p^n$ and integers $\ell^{(1)}, \ldots, \ell^{(k)} \leq p-1$ such that*

$$|L(P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})})| \geq \frac{n}{dp^{\lceil \frac{d-1}{p-1} \rceil}}.$$

Proof. The exists some $d' \leq d$ so that $\deg_i(P) = d'$ for at least $\frac{n}{d}$ variables, call this set of variables G. If $d' = 1$, then the claim holds trivially, so assume $d' > 1$. Let $d' - 1 = (k-1)(p-1) + \ell$ for $\ell \leq p-2$ and set $\ell^{(1)} = \cdots = \ell^{(k-1)} = p-1, \ell^{(k)} = \ell$. Then applying Lemma 3.2, for every $x_i \in G$, the coefficient $c_i(y^{(1)}, \ldots, y^{(k)})$ of x_i in $P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}$ is a non-zero polynomial of degree at most $d' - 1 \leq d - 1$. Thus, there exists a setting for $y_1, \ldots, y_k$ where at least

$$\delta_p(d-1)|G| \geq \frac{n}{dp^{\lceil \frac{d-1}{p-1} \rceil}}$$

of the c_is are non-zero. Since variables in G have degree 1 in $P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}$, there are no higher degree terms which contain them, so these variables all lie in $L(P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})})$. □

To complete the proof of Lemma 3.1, we observe that $P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}$ can be written as

$$P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}(x) = \sum_{i \leq t} \lambda_i P(x + a_i)$$

$$\text{where } \; t \leq \prod_{j=1}^{k} (\ell^{(j)} + 1) \leq p^{\lceil \frac{d-1}{p-1} \rceil}.$$

4 The case of characteristic 2

Let $P(x)$ be a low degree polynomial over $\mathbb{Z}_2$. We prove in this section that P must have high degree over characteristics $q \neq 2$. Since we will be working with operations over different fields, we will use $+_p$ to denote summation modulo p, and $\oplus$ for summation modulo 2. We start with the some simple claims:

Claim 4.1. *Let $f(x) = \oplus_{i=1}^n x_i$ be the parity function on n bits. Then for $q \neq 2$, $\deg_q(f) = n$.*

Proof. The unique multilinear polynomial over $\mathbb{Z}_q$ computing f is

$$H^{\oplus}(x) = \frac{1}{2}(1 - \prod_{i=1}^{n}(1 - 2x_i))$$

□

Lemma 4.2. *Let $a_1, \ldots, a_k \in \mathbb{Z}_2^n$. Define $g : \{0,1\}^n \to \{0,1\}$ by $g(x) = \oplus_{i=1}^k f(x \oplus a_i)$. Then*

$$\deg_q(g) \leq k \deg_q(f)$$

Proof. For any $a \in \mathbb{Z}_2^n$, consider $f_a(x) = f(x \oplus a)$. We claim that $\deg_q(f_a) = \deg_q(f)$. Let $Q(x)$ be a polynomial over $\mathbb{Z}_q$ which computes f over $\{0,1\}^n$. Define a new polynomial $Q_a(x) = Q(x \oplus a)$ by replacing x_i with $1 - x_i$ whenever $a_i = 1$, and keeping x_i whenever $a_i = 0$. Clearly Q_a computes $f_a(x)$ over $\{0,1\}^n$, and $\deg_q(Q_a) = \deg_q(Q)$. Note that $g(x) = \oplus_{i=1}^k f_{a_i}(x)$.

Composing the polynomial $H^{\oplus}$ over $\mathbb{Z}_q$ that computes $\oplus$ on $\{0,1\}^k$ with the Q_as, we get a polynomial of degree at most $k \deg_q(f)$ that represents g over $\mathbb{Z}_q$, thus $\deg_q(g) \leq k \deg_q(f)$. □

We now restate and prove Theorem 1.2 in the $p = 2$ case, showing that any Boolean function with small degree over $\mathbb{Z}_2$ must have high degree over $\mathbb{Z}_q$ for a prime $q \neq 2$.

Theorem 4.3 (Theorem 1.2, $p = 2$ case)**.** *For any $f : \{0,1\}^n \to \{0,1\}$, and prime $q \neq 2$:*

$$\deg_q(f) \geq \frac{n}{\deg_2(f) 4^{\deg_2(f)}}.$$

Proof. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, let $\deg_2(f) = d$. Let $P(x)$ be the degree d polynomial over $\mathbb{Z}_2$ computing f. We will prove that the multilinear polynomial $Q(x)$ over $\mathbb{Z}_q$ computing f has high degree.

By Lemma 3.1, there exist $a_1, \ldots, a_k \in \mathbb{Z}_2^n$ where $k \leq 2^d$, such that if $P'(x) = \oplus_{i=1}^k P(x + a_i)$, then $|L(P')| \geq \frac{n}{d2^d}$. Let us denote the set $L(P')$ by S. Let P'_S be the restriction of P' to the variables in S obtained by fixing the remaining variables to zero; $P'_S(x)$ is either Parity on the set S or its negation, assume w.l.o.g it is the former.

Now consider the polynomial Q. Since $Q(x) = f(x)$ for all $x \in \{0,1\}^n$, then the polynomial Q' defined as $Q'(x) = H^{\oplus}(Q(x \oplus a_1), \ldots, Q(x \oplus a_k))$ satisfies that $Q'(x) = P'(x)$ for all $x \in \{0,1\}^n$. So if we let Q'_S be the restriction of Q' to the variables in S, then $Q'_S(x) = P'_S(x)$ for all $x \in \{0,1\}^n$.

Now, since P'_S is the parity function over $|S|$ bits, Claim 4.1 implies that $\deg(Q'_S) = |S| \geq \frac{n}{d2^d}$. On the other hand, we have $\deg(Q'_S) \leq \deg(Q') \leq k \deg_q(f)$ by Lemma 4.2. We conclude that

$$\deg_q(f) \geq \frac{n}{kd2^d} \geq \frac{n}{d4^d}$$

□

We now generalize this result and show that f cannot be approximated by low degree polynomials over $\mathbb{Z}_q$. We need the following claim, which is proven using the union bound.

Claim 4.1. *Let $f' : \{0,1\}^n \to \{0,1\}$ be such that $\Pr_{x \in \{0,1\}^n}[f'(x) = f(x)] \geq 1 - \epsilon$. Let $a_1, \ldots, a_k \in \mathbb{Z}_2^n$. Then*

$$\Pr_{x \in \{0,1\}^n}[\oplus_{i=1}^k f'(x \oplus a_i) = \oplus_{i=1}^k f(x \oplus a_i)] \geq 1 - k\epsilon.$$

We now restate and prove Theorem 1.4 in the $p = 2$ case.

Theorem 4.4 (Theorem 1.4, $p = 2$ case)**.** *For prime $q \neq 2$ let $c, \epsilon > 0$ be given by Lemma 2.2. Let $f : \{0,1\}^n \to \{0,1\}$ and $\deg_2(f) = d$. If $f' : \{0,1\}^n \to \{0,1\}$ satisfies*

$$\Pr_{x \in \{0,1\}^n}[f'(x) = f(x)] \geq 1 - 2^{-d}\epsilon,$$

then

$$\deg_q(f') \geq c\sqrt{\frac{n}{d8^d}}.$$

Proof. Using Lemma 3.1, choose $k \leq 2^d$ and $a_1, \ldots, a_k \in \mathbb{Z}_2^n$ so that $g(x) = \oplus_{i=1}^k f(x \oplus a_i)$ when restricted to a set S is either Parity or its negation on $|S| \geq \frac{n}{d2^d}$ variables.

Define $g'(x) = \oplus_{i=1}^k f'(x \oplus a_k)$. By Claim 4.1 we get that

$$\Pr_{x \in \{0,1\}^n}[g(x) = g'(x)] \geq 1 - k2^{-d}\epsilon \geq 1 - \epsilon.$$

For every assignment $b \in \{0,1\}^{[n] \setminus S}$ to the variables outside S, define $g_{S,b}(x)$ be the restriction of g to the variables in S, obtained by assigning the values of the variables outside S according to b. Similarly define $g'_{S,b}$. We claim there exists some b such that

$$\Pr_{x \in \{0,1\}^S}[g_{S,b}(x) = g'_{S,b}(x)] \geq 1 - \epsilon.$$

Indeed, this is true as for a randomly chosen b,

$$\begin{aligned}&\mathbf{E}_{b \in \{0,1\}^{[n] \setminus S}}\left[\Pr_{x \in \{0,1\}^S}[g_{S,b}(x) = g'_{S,b}(x)]\right] \\ &= \Pr_{x \in \{0,1\}^n}[g(x) = g'(x)] \geq 1 - \epsilon.\end{aligned}$$

We also have $\deg_q(g'_{S,b}) \leq \deg_q(g') \leq 2^d \deg_q(f')$, where the last inequality uses Lemma 4.2. Now, $g_{S,b}(x)$ is either Parity or its negation (assume w.l.o.g the former) over $|S|$ variables. So $g'_{S,b}$ has degree at most $2^d \deg_q(f')$ and approximates Parity over $|S|$ variables with probability at least $1 - \epsilon$. By Lemma 2.2 this implies $\deg_q(g'_{S,b}) \geq c\sqrt{|S|}$.Thus

$$2^d \deg_q(f') \geq \deg(g'_{S,b}) \geq c\sqrt{\frac{n}{d2^d}}$$

which proves the theorem. □

Combining Theorem 4.4 with the Razborov-Smolensky bound, we conclude that any $AC_0[q]$ circuit that computes a low $\mathbb{Z}_2$-degree Boolean function on n variables must be of exponential size.

Theorem 4.5 (Theorem 1.5, $p = 2$ case)**.** *For any prime $q \neq 2$, there exist a constant c_1 so that any $AC_0[q]$ circuit of depth t computing a function $f : \{0,1\}^n \to \{0,1\}$ on n variables with $\deg_2(f) = d$ requires size $c_1 2^{-d} \exp((\frac{n}{d8^d})^{\frac{1}{2t}})$.*

Proof. Assume there is an $AC_0[q]$ circuit of size s and depth t computing f. Let ϵ be the constant in Lemma 2.2. Applying Lemma 2.1 with $\delta = 2^{-d}\epsilon$, there is some universal constant c' and an $\mathbb{F}_q$ polynomial Q of degree $\deg(Q) \leq \left(c' \log \frac{s}{2^{-d}\epsilon}\right)^t$ such that

$$\Pr_{x \in \{0,1\}^n}[Q(x) = f(x)] \geq 1 - 2^{-d}\epsilon.$$

By Theorem 4.4 we get that $\deg(Q) \geq c\sqrt{\frac{n}{d8^d}}$ for some constant c. Hence,

$$s \geq c_1 2^{-d} \exp\left(\left(\frac{n}{d8^d}\right)^{\frac{1}{2t}}\right),$$

for a universal constant c_1. □

5 The case of general characteristic

Since we will be working with operations over different fields, we will denote by $+_p, +_q$ summation modulo p, q respectively, and by $+$ summation where the context is clear.

Mapping $\mathbb{Z}_p^n$ into $\mathbb{Z}_q^{n'}$: Let $f(x)$ be a Boolean function. We start by defining a polynomial extending this into a function $F : \mathbb{Z}_p^n \to \{0,1\}$. Given a vector $x \in \mathbb{Z}_p^n$, we define $x^{p-1} = (x_1^{p-1}, \ldots, x_n^{p-1}) \in \{0,1\}^n$, which is the indicator of whether x is non-zero on each coordinate. Define the function $F : \mathbb{Z}_p^n \to \{0,1\}$ by $F(x) = f(x^{p-1})$. $F(x)$ can be expressed as a polynomial of degree $(p-1)\deg_p(f)$ by taking the multilinear representation of f over $\mathbb{Z}_p$ and replacing each variable x_i with x_i^{p-1}; henceforth we think of F as this polynomial. Our goal will be to show that if f has low degree over $\mathbb{Z}_q$, then so does F and any function of the form $F(x +_p a_1) +_p \ldots +_p F(x +_p a_k)$. Since these are functions on $\mathbb{Z}_p^n$, we need to define the notion of computing functions on $\mathbb{Z}_p^n$ by polynomials over $\mathbb{Z}_q$. Set $b = \lceil \log_2 p \rceil$. We identify the lexicographically first p bit strings in $\{0,1\}^b$ with the set $\{0, \ldots, p-1\}$. We then identify $\mathbb{Z}_p^n$ with a subset of $\mathbb{Z}_q^{nb}$ by identifying $x = (x_1, \ldots, x_n) \in \mathbb{Z}_p^n$ with $(x_{1,1}, \ldots, x_{1,b}, \ldots, x_{n,1}, \ldots, x_{n,b}) \in \mathbb{Z}_q^{nb}$, where the value of x_i determines the values of $(x_{i,1}, \ldots, x_{i,b})$. Notice that in fact we map $\mathbb{Z}_p^n$ into $\{0,1\}^{nb} \subset \mathbb{Z}_q^{nb}$. Given $x \in \mathbb{Z}_p^n$, we use $\bar{x} \in \{0,1\}^{nb}$ to denote the vector in $\{0,1\}^{nb} \subset \mathbb{Z}_q^{nb}$ that represents it. We use $\bar{x_i}$ to denote the vector $(x_{i,1}, \ldots, x_{i,b})$ that represents x_i. We say a polynomial $G(x) \in \mathbb{Z}_q[x_{1,1}, \ldots, x_{n,b}]$ computes $F : \mathbb{Z}_p^n \to \{0,1\}$ if $F(x) = G(\bar{x})$ for every $x \in \mathbb{Z}_p^n$.

We start by showing that if f has low degree in $\mathbb{Z}_q$, then $F(x +_p a)$ can also be computed by a low degree polynomial over $\mathbb{Z}_q$.

Lemma 5.1. *Let $f : \{0,1\}^n \to \{0,1\}$ and let $F(x)$ be a polynomial over $\mathbb{Z}_p$ defined by $F(x) = f(x^{p-1})$. For every $a \in \mathbb{Z}_p^n$ there is a polynomial $G_a(x) \in \mathbb{Z}_q[x_{1,1}, \ldots, x_{n,b}]$ over $\mathbb{Z}_q$ of degree at most $b\deg_q(f)$ that computes $F(x +_p a)$.*

Proof. Given $a = (a_1, \ldots, a_n) \in \mathbb{Z}_p^n$, we can define $A_i(\bar{x_i}) \in \mathbb{Z}_q[\bar{x_i}]$ for every $i \in [n]$ so that $\deg(A_i) \leq b$ and

$$A_i(\bar{x_i}) = \begin{cases} 0 & \text{if } x_i + a_i = 0 \bmod p \\ 1 & \text{otherwise} \end{cases}$$

It follows that $(A_1(\bar{x_1}), \ldots, A_n(\bar{x_n})) = (x +_p a)^{p-1}$. We now define the polynomial $G_a(\bar{x})$ as

$$G_a(\bar{x}) = F_{\{0,1\}}(A_1(\bar{x_1}), \ldots, A_n(\bar{x_n}))$$

where $F_{\{0,1\}}$ is the multilinear polynomial over $\mathbb{Z}_q$ computing f over $\{0,1\}^n$. We have:

$$G_a(\bar{x}) = F_{\{0,1\}}((x +_p a)^{p-1}) = f((x +_p a)^{p-1}) = F(x +_p a)$$

as required, and $\deg(G_a) \leq b \deg(F_{\{0,1\}}) = b \deg_q(f)$. □

Our goal will be to compute Boolean predicates on sums $F(x +_p a_1) +_p \ldots +_p F(x +_p a_k)$ by low degree polynomials over $\mathbb{Z}_q$.

Corollary 5.2. *Let $f : \{0,1\}^n \to \{0,1\}$ and let $F(x)$ be a polynomial over $\mathbb{Z}_p$ defined by $F(x) = f(x^{p-1})$. Let $a_1, \ldots, a_k \in \mathbb{Z}_p^n$ be points, and let $t : \mathbb{Z}_p \to \{0,1\}$ be any Boolean valued predicate on $\mathbb{Z}_p$. Define the function $T : \mathbb{Z}_p^n \to \{0,1\}$ to be*

$$T(x) = t(\sum_{i \leq k} \lambda_i F(x +_p a_i))$$

Then, T can be computed by a polynomial over $\mathbb{Z}_q$ of degree at most $kb \deg_q(f)$.

Proof. By Lemma 5.1, each function $F(x +_p a_i)$ can be computed by a polynomial $G_i(\bar{x})$ over $\mathbb{Z}_q$ of degree at most $b \deg_q(f)$. The function $T(x)$ is a function of $G_1(\bar{x}), \ldots, G_k(\bar{x}) \in \{0,1\}$, and thus can be computed by $H(G_1(\bar{x}), \ldots, G_k(\bar{x}))$, where $H(z_1, \ldots, z_k)$ is a multilinear polynomial over $\mathbb{Z}_q$ computing the function $t(\lambda_1 z_1 +_p \ldots +_p \lambda_k z_k) : \{0,1\}^k \to \{0,1\}$. Thus, T can be computed by a polynomial over $\mathbb{Z}_q$ of degree at most $kb \deg_q(f)$. □

We now prove Theorem 1.2 in the case of general p.

Proof of Theorem 1.2 for general p. Let $d = \deg_p(f)$, and consider $F(x) = f(x^{p-1})$ which has degree $(p-1)d$. Invoking Lemma 3.1 for $F(x)$ which has degree $(p-1)d$, we conclude that there exist $k \leq p^d$ points $a_1, \ldots, a_k \in \mathbb{Z}_p^n$ such that $G(x) = \sum_{i=1}^k \lambda_i F(x +_p a_i)$ satisfies $|L(G)| \geq n/(dp^d)$. Let $S = L(G)$, and rename the variables in S as $x_1, \ldots, x_s$, where $s = |S|$. If we let G_S be the restriction of G to variables in S (by setting the other variables to zero), we have

$$G_S(x) = \sum_{i \leq s} \lambda_i x_i + c, \quad \lambda_i \in \mathbb{Z}_p \setminus \{0\}, c \in \mathbb{Z}_p.$$

Let ω be a p^{th} root of unity in the appropriate extension field $\mathbb{F} = \mathbb{F}_{q^h}$ of $\mathbb{F}_q$. We consider the function $h : \{0,1\}^s \to \mathbb{F}$ given by $h(x) = \omega^{\sum_{i \leq s} \lambda_i x_i + c}$. The unique multilinear polynomial $H(x)$ over $\mathbb{F}$ computing h over $\{0,1\}^s$ has degree $\deg_{\mathbb{F}}(H) = s \geq \frac{n}{dp^d}$ and is given by

$$H(x) = \omega^c \prod_{i=1}^{s} (1 + (\omega^{\lambda_i} - 1)x_i)$$

But we can upper-bound $\deg(H)$ in terms of $\deg_q(f)$. First, for $i \in \{0, \cdots, p-1\}$ let $t_i : \mathbb{Z}_p \to \{0,1\}$ be the predicate indicating whether $x \equiv i \bmod p$. We can obtain the polynomial $H(x)$ by multlinearization of the polynomial

$$H'(x) = \sum_{i=0}^{p-1} \omega^i t_i(G_S(x))$$

Since $G_S(x)$ is of the form $\sum_i \lambda_i F(x +_p a_i)$, Corollary 5.2 gives $\deg_q(t_i(G_S(x))) \leq kb \deg_q(f)$. Hence,

$$\deg_{\mathbb{F}}(H) \leq \max_i \deg_q(t_i(G_S(x))) \leq kb \deg_q(f).$$

Thus we get

$$\deg_q(f) \geq \frac{s}{bk} \geq \frac{n}{\lceil \log_2 p \rceil dp^{2d}}.$$

□

Next we prove Theorem 1.4, that functions with low degree over $\mathbb{Z}_p$ are hard to approximate over $\mathbb{Z}_q$. First we state the theorem precisely.

Theorem 5.3 (Theorem 1.4 for general p)**.** *For prime $q \neq p$ let $c, \epsilon > 0$ be given by Lemma 2.2. Let f be a Boolean function such that $\deg_p(f) = d$. Let f' be any Boolean function satisfying*

$$\Pr_{x \in \{0,1\}^n} [f'(x) = f(x)] \geq 1 - p^{-d}\epsilon.$$

Then

$$\deg_q(f') \geq \frac{c}{\lceil \log_2 p \rceil} \sqrt{\frac{n}{dp^{3d}}}$$

We start with some technical claims.

Claim 5.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, such that $\deg_p(f) = d$. For $v \in \{0,1\}^n$ define $F_v : \mathbb{Z}_p^n \to \{0,1\}$ as*

$$F_v(x) = f(x^{p-1} \oplus v)$$

where for $y, v \in \{0,1\}^n$, $y \oplus v \in \{0,1\}^n$ denotes their coordinatewise-Xor. Then F_v is a polynomial over $\mathbb{Z}_p$ of degree at most $(p-1)d$.

To prove this claim, we construct the polynomial for F_v from the multilinear polynomial for f by replacing x_i with x^{p-1} or $1 - x^{p-1}$ depending on whether or not $v_i = 0$. We omit the details.

Claim 5.5. *Let $f(x)$ and $f'(x)$ be two Boolean functions such that*

$$\Pr_{x \in \{0,1\}^n}[f(x) = f'(x)] \geq 1 - \epsilon.$$

There exists $v \in \{0,1\}^n$ such that if we define $F_v(x) = f(x^{p-1} \oplus v)$ and $F'_v = f'(x^{p-1} \oplus v)$ then

$$\Pr_{x \in \mathbb{Z}_p^n}[F_v(x) = F'_v(x)] \geq 1 - \epsilon.$$

Proof. If we choose $v \in \{0,1\}^n$ at random, then

$$\mathbf{E}_v[\Pr_{x \in \mathbb{Z}_p^n}[F_v(x) = F'_v(x)]] = \Pr_{x \in \{0,1\}^n}[f(x) = f'(x)] \geq 1 - \epsilon.$$

Thus the inequality holds for some $v \in \{0,1\}^n$, □

We also need the following analogue of Claim 4.1:

Claim 5.6. *Let $F(x)$ and $F'(x)$ be functions such that $\Pr_{x \in \mathbb{Z}_p^n}[F(x) = F'(x)] \geq 1 - \epsilon$. Let $a_1, \ldots, a_k \in \mathbb{Z}_p^n$ and $\lambda_1, \ldots, \lambda_k \in \mathbb{Z}_p$. Then:*

$$\Pr_{x \in \mathbb{Z}_p^n}[\sum_i \lambda_i F(x +_p a_i) = \sum_i \lambda_i F'(x +_p a_i)] \geq 1 - k\epsilon.$$

We now prove Theorem 5.3.

Proof of Theorem 1.4 in the case of general p. Let $f(x)$ be a Boolean function of small degree d over $\mathbb{Z}_p$. Let $f'(x)$ be another Boolean function such that $\Pr_{x \in \{0,1\}^n}[f(x) = f'(x)] \geq 1 - p^{-d}\epsilon$. We will prove that $\deg_q(f')$ is large. The proof will proceed by a series of transformations on the pair of functions, such that the pairs generated will remain close, f will be transformed into the Mod_p function, whereas f' will be transformed into a function whose degree over $\mathbb{Z}_q$ can be bounded by a function of $\deg_q(f')$.

The first step is to extend f, f' to functions mapping $\mathbb{Z}_p^n$ to $\{0,1\}$. By Claim 5.5, there exists $v \in \{0,1\}^n$ such that

$$\Pr_{x \in \mathbb{Z}_p^n}[F_v(x) = F'_v(x)] \geq \Pr_{x \in \{0,1\}^n}[f(x) = f'(x)] \geq 1 - p^{-d}\epsilon.$$

Also, the degree of F_v over $\mathbb{Z}_p$ is at most $(p-1)d$.

The next step is to apply the degree reduction lemma to F_v. By Lemma 3.1, there exists $k \leq p^d$ and points $a_1, \ldots, a_k \in \mathbb{Z}_p^n$, such that if $G(x) = \sum_{i \leq k} \lambda_i F_v(x +_p a_i)$ (the sum is addition modulo p), then the set $S = L(G)$ will have size $s \geq \frac{n}{dp^d}$. If we define $G' : \mathbb{Z}_p^n \to \mathbb{Z}_p$ as

$$G'(x) = \sum_{i \leq k} \lambda_i F'_v(x +_p a_i) \tag{4}$$

then Claim 5.6 implies

$$\Pr_{x \in \mathbb{Z}_p^n}[G(x) = G'(x)] \geq 1 - kp^{-d}\epsilon \geq 1 - \epsilon.$$

As in the proof of Theorem 4.4, there exists an assignment $u \in \mathbb{Z}_p^{[n]\setminus S}$ to the variables outside S so that the agreement between G and G' is at least as large. To ease notation, we denote these restrictions as $G(x)$ and $G'(x)$ (as opposed to $G_{S,u}(x)$ and $G'_{S,u}(x)$). Note that $G(x) = \sum_{i \leq k} \lambda_i x_i + c$ where $\lambda_i \in \mathbb{Z}_p \setminus \{0\}$, $c \in \mathbb{Z}_p$ and the summation is modulo p. By replacing each x_i in G by $\lambda_i^{-1} x_i$, we get a new function $G'' : \mathbb{Z}_p^s \to \mathbb{Z}_p$ where

$$\Pr_{x \in \mathbb{Z}_p^s}[G''(x) = \sum_i x_i + c] \geq 1 - \epsilon.$$

The final step is to get a function h on $\{0,1\}^n$ which computes the Mod_p function on s variables. Towards this, for each $w \in \mathbb{Z}_p^s$, we define $g_w : \{0,1\}^s \to \mathbb{Z}_p$ by $g_w(y) = G''(y + w)$. Note that we have:

$$\begin{aligned} &\Pr_{w \in \mathbb{Z}_p^s}[\Pr_{y \in \{0,1\}^s}[g_w(y) = \sum_i y_i + \sum_i w_i + c]] \\ &= \Pr_{x \in \mathbb{Z}_p^s}[G''(x) = \sum_i x_i + c] \geq 1 - \epsilon \end{aligned}$$

since $y + w$ is distributed uniformly at random over $\mathbb{Z}_p^s$. Thus there exists a good w so that:

$$\begin{aligned} &\Pr_{y \in \{0,1\}^s}[g_w(y) = \sum_{i \leq s} y_i + c'] \geq 1 - \epsilon \\ &\text{where } \ c' = c + \sum_i w_i \in \mathbb{Z}_p. \end{aligned}$$

Define $t : \mathbb{Z}_p \to \{0,1\}$ by $t(z) = 1$ iff $z \equiv c' \bmod p$ and $t(z) = 0$ otherwise. Finally, let $h(y) = t(g_w(y))$. We have

$$\Pr_{y \in \{0,1\}^s}[h(y) = \mathsf{Mod}_p(y)] \geq 1 - \epsilon.$$

Hence Lemma 2.1 implies that $\deg_q(h) \geq c\sqrt{s}$.

Our goal now is to relate $\deg_q(h)$ to $\deg_q(f')$. We make the following observations:

1. We have $g_w(y) = G''(y + w)$.

2. $G''(x)$ is obtained from $G'(x)$ by setting variables outside S to constants and replacing each $x_i \in S$ by $\lambda^{-1}x_i$.

3. By Equation 4, $G'(x)$ is a linear combination of values of the form $F'_v(x +_p a_i)$.

4. Each $F'_v(x +_p a_i)$ can be computed by a polynomial $Q_i(\bar{x})$ over $\mathbb{Z}_q$ of degree at most $b \deg_q(f')$ by an argument similar to Lemma 5.1.

Thus, we can write $h(y)$ as some predicate $t' : \{0,1\}^k \to \{0,1\}$ applied to a tuple of polynomial $Q_1, \dots, Q_k$ with $\deg_q(Q_i) \le b \deg_q(f')$, and hence $\deg_q(h) \le kb \deg_q(f')$.

We conclude that

$$\deg_q(f') \ge \frac{c\sqrt{s}}{kb} = \frac{c}{\lceil \log_2 p \rceil}\sqrt{\frac{n}{dp^{3d}}}.$$

□

As a corollary, we get a lower bound for the size of $AC_0[q]$ circuits computing functions with low degree over $\mathbb{Z}_p$:

Theorem 5.7 (Theorem 1.5, restated). *Let p, q be distinct primes. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function on n variables with $\deg_p(f) = d$. Then any $AC_0[q]$ circuit of depth t computing f requires size at least*

$$c_1 p^{-d} \exp\left(\left(\frac{n}{\lceil \log_2 p \rceil^2 dp^{3d}}\right)^{\frac{1}{2t}}\right),$$

where c_1 is a universal constant. In particular, for $d = o(\log_p n)$, the lower bound is $\exp(n^{1/2t - o(1)})$.

Proof. Assume there is an $AC_0[q]$ circuit of size s and depth t computing f. Let ϵ be the constant in Lemma 2.2. Applying Lemma 2.1 with $\delta = p^{-d}\epsilon$ implies that there is some universal constant c' and an $\mathbb{F}_q$ polynomial Q of degree $\deg(Q) \le \left(c' p \log \frac{s}{p^{-d}\epsilon}\right)^t$ such that $\Pr_{x \in \{0,1\}^n}[Q(x) = f(x)] \ge 1 - p^{-d}\epsilon$. By Theorem 5.3 $\deg(Q) \ge c\sqrt{\frac{n}{\lceil \log_2 p \rceil^2 dp^{3d}}}$ for some c. Hence, there is a constant c_1 so that

$$s \ge c_1 p^{-d} \exp\left(\left(\frac{n}{\lceil \log_2 p \rceil^2 dp^{3d}}\right)^{\frac{1}{2t}}\right),$$

□

References

[ABFR94] J. Aspnes, R. Beigel, M. Furst, and S. Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):1–14, 1994.

[BBR94] David A. Barrington, Richard Beigel, and Steven Rudich. Representing Boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4:367–382, 1994.

[Bei93] Richard Beigel. The polynomial method in circuit complexity. *Structures in Complexity Theory: 8^{th} Annual Conference*, pages 82–95, 1993.

[BGL06] Nayantara Bhatnagar, Parikshit Gopalan, and Richard J. Lipton. Symmetric polynomials over $\mathbb{Z}_m$ and simultaneous communication protocols. *Journal of Computer and System Sciences*, 72:252–285, 2006.

[BRS91] R. Beigel, N. Reingold, and D. Spielman. The perceptron strikes back. In *Proceedings of the Sixth Conference on Structure in Complexity Theory*, pages 286–291, 1991.

[BV07] Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. In *48^{th} Annual Symposium on Foundations of Computer Science (FOCS'07)*, pages 41–51. IEEE, 2007.

[Cho61] C.K. Chow. On the characterization of threshold functions. In *Proceedings of the Symposium on Switching Circuit Theory and Logical Design (FOCS)*, pages 34–38, 1961.

[Efr09] Klim Efremenko. 3-query locally decodable codes of exponential codes. In *Accpeted to the 41^{st} Annual Symposium on the Theory of Computing (STOC'09).* ACM, 2009.

[Gop06a] Parikshit Gopalan. *Computing with Polynomials over Composites.* PhD thesis, Georgia Institute of Technology, 2006.

[Gop06b] Parikshit Gopalan. Constructing Ramsey graphs from Boolean function representations. In *Proceedings of the 21^{st} IEEE Conference on Computational Complexity (CCC'06)*, 2006.

[Gro00] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20(1):71–86, 2000.

[Gro02] Vince Grolmusz. Constructing set systems with prescribed intersection sizes. *Journal of Algorithms*, 44(2):321–337, 2002.

[JKS02] J. Jackson, A. Klivans, and R. Servedio. Learnability beyond AC^0. In *Proceedings of the 34th ACM Symposium on Theory of Computing*, 2002.

[KKMS05] A. T. Kalai, A. R. Klivans, Y. Mansour, and R. Servedio. Agnostically learning halfspaces. In *Proc. 46^{th} IEEE Symp. on Foundations of Computer Science (FOCS'05)*, 2005.

[KM93] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. on Computing*, 22(6):1331–1348, 1993.

[KOS02] A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. In *Proceedings of the* 43^{rd} *Annual Symposium on Foundations of Computer Science (FOCS'02)*, pages 177–186, 2002.

[KS01] A. Klivans and R. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. In *Proceedings of the* 33^{rd} *Annual Symposium on Theory of Computing (STOC'01)*, pages 258–265, 2001.

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, Fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993.

[Lov08] Shachar Lovett. Unconditional pseudorandom generators for low degree polynomials. In 40^{th} *Annual Symposium on the Theory of Computing (STOC'08)*, pages 557–562. ACM, 2008.

[MOS03] Elchannan Mossel, Ryan O'Donnell, and Rocco Servedio. Learning juntas. In *Proceedings of the* 35^{th} *Annual ACM Symposium on the Theory of Computing (STOC'03).*, 2003.

[MP68] Marvin Minsky and Seymour Papert. *Perceptrons: an Introduction to Computational Geometry.* MIT Press, 1968.

[MS77] F. MacWilliams and N. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, 1977.

[Mur71] S. Muroga. *Threshold logic and its applications.* Wiley-Interscience, New York, 1971.

[NS92] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. In *Proceedings of the* 24^{th} *Annual ACM Symposium on the Theory of Computing (STOC'92)*, pages 462–467, 1992.

[Pat92] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proceedings of the 24th Symposium on Theory of Computing*, pages 468–474, 1992.

[Raz87] Alexander Razborov. Lower bounds for the size of circuits of bounded depth with basis $\{\wedge, \oplus\}$. *Methematical Notes of the Academy of Science of the USSR*, 41:333–338, 1987.

[Smo87] Roman Smolensky. Algebraic methods in the theory of lower bounds for Boolean circuit complexity. In *Proceedings of the* 19^{th} *Annual ACM Symposium on Theoretical Computer Science (STOC'87)*, pages 77–82, 1987.

[TB98] Gabor Tardos and David Barrington. A lower bound on the mod 6 degree of the OR function. *Computational Complexity*, 7:99–108, 1998.

[Vio08] Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. In *Proceedings of the* 23^{rd} *IEEE Conference on Computational Complexity (CCC'08)*, 2008.

ECCC ISSN 1433-8092

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