comput. complex. **19** (2010), 235–263
1016-3328/10/020235-29, *published online* 12 May 2010
DOI 10.1007/s00037-010-0290-4

computational complexity

THE COMPLEXITY OF BOOLEAN FUNCTIONS IN DIFFERENT CHARACTERISTICS

Parikshit Gopalan, Shachar Lovett,
and Amir Shpilka

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 at most 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, then 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 $\mathrm{AC}_0[q]$ circuits. This gives a lower bound for a broad and natural class of functions.

Keywords. Low degree polynomials, bounded depth circuits, lower bounds.

Subject classification. 68Q17.

1. Introduction

Representations of Boolean functions as polynomials in various characteristics have been studied intensively in Computer science (Barrington *et al.* 1994;

Beigel 1993; Nisan & Szegedy 1992; Paturi 1992). This algebraic view of Boolean functions has found numerous applications to diverse areas including circuit lower bounds (Aspnes *et al.* 1994; Beigel *et al.* 1991; Razborov 1987; Smolensky 1987), computational learning (Klivans & Servedio 2001; Kushilevitz & Mansour 1993; Linial *et al.* 1993; Mossel *et al.* 2003) and explicit combinatorial constructions (Efremenko 2009; Gopalan 2006b; Grolmusz 2000, 2002). 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 $k = p$ ($R = \mathbb{F}_p$) 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 k, we have

$$\deg_k(f) \leq \deg(f)\,.$$

This is because the polynomial representing f over $\mathbb{Z}/k\mathbb{Z}$ can be obtained from the representation over $\mathbb{Z}$ by taking each coefficient modulo k. 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 prime $p \neq 2$.

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 $\mathrm{AC}_0[q]$ 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 Nisan & Szegedy 1992) 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 showed that any function on n variables must have degree at least $\deg(f) \geq \log n - O(\log\log n)$ (Nisan & Szegedy 1992). An interesting consequence of Theorem 1.2 is the following analog 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 m is not a prime power, and p is its smallest prime divisor. We have*

$$\deg_m(f) \geq \frac{1}{2}\log_p n - \log_p \log_p n - \frac{1}{2}\log_p \lceil \log_2 p \rceil .$$

This corollary is interesting as it illuminates a sharp difference between degrees over composite numbers and over primes. A simple way to construct Boolean functions of 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 analog 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 immediately implies a lower bound for $\deg(f)$, one can obtain the following stronger bound by a simple modification of the Nisan–Szegedy proof:

LEMMA 1.4. *Let p be a prime and $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which depends on all n variables. Then*

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

We prove this lemma in Section 2.1.

The results above show a very basic relation between the degrees of Boolean functions over different characteristics. 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 Smolensky (1987), we prove that low degree polynomials modulo p are hard to even approximate by polynomials in other characteristics.

THEOREM 1.5. *Let $f : \{0,1\}^n \to \{0,1\}$ be a function depending on all n variables with $\deg_p(f) = d$. Then, for any $q \neq p$ and any $\mathbb{F}_q$ polynomial $Q(x_1, \dots, x_n) : \mathbb{F}_q^n \to \{0,1\}$, satisfying $\deg_q(Q) = o(\sqrt{\frac{n}{dp^{3d}}})$, it holds that*

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

where ϵ depends only on p, 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 $\mathbb{F}_p$ that are 0 on the boolean hypercube with probability $1 - 2^{-d}$, hence they have trivial approximations over $\mathbb{F}_q$. Secondly, the Mod_p function (and indeed every symmetric function) can be $1 - \epsilon$ approximated by polynomials of degree $c(\epsilon)\sqrt{n}$ over $\mathbb{F}_q$ (Bhatnagar *et al.* 2006), despite being hard to approximate for polynomials of lower degree.

As a corollary of Theorem 1.5 we get that if a Boolean function has low degree modulo p, then the function requires large $\mathrm{AC}_0[q]$ circuits for any prime $q \neq p$. Several of the known lower bounds for $\mathrm{AC}_0[q]$ are for 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 $\mathrm{AC}_0[q]$, namely all functions that have degree $o(\log n)$ modulo $p \neq q$.

THEOREM 1.6. *Let p, q be distinct primes. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function which depends on all n variables with $\deg_p(f) = o(\log_p n)$. Then any $\mathrm{AC}_0[q]$ 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 Smolensky (1987) follows from the observation that $\deg_p(\mathsf{Mod}_{p^k}) \leq p^k$ (see e.g.

Bhatnagar *et al.* 2006). Additionally, it gives several new lower bounds, for instance it shows that every quadratic form on n variables over $\mathbb{F}_2$ requires large $\mathrm{AC}_0[q]$ circuits, for $q \neq 2$. Though we note that Theorem 1.6 does not imply Razborov's lower bound for $\mathsf{Majority}$.

Summarizing, Theorems 1.2 and 1.5 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 (Muroga 1971), voting theory (Chow 1961) and machine learning (Minsky & Papert 1968). 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 MacWilliams & Sloane (1977, Chapters 13-14) and the references therein. The codewords of the code $\mathsf{RM}_2(d, n)$ 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 (Beigel 1993) 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 $\mathrm{AC}_0[p]$ (Razborov 1987; Smolensky 1987), which relates the circuit size to the polynomial degree needed to approximate f over $\mathbb{F}_p$, and the work of Beigel *et al.* (1991) and Aspnes *et al.* (1994) 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 (Kushilevitz & Mansour 1993), DNF formulae (Klivans & Servedio 2001), AC_0 circuits (Jackson *et al.* 2002; Linial *et al.* 1993), juntas (Mossel *et al.* 2003) and halfspaces (Kalai *et al.* 2005; Klivans *et al.* 2002) all proceed by showing that the concept class to be learned has some *nice* polynomial representation. In particular, the algorithm for learning juntas of Mossel *et al.* (2003) 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 (Grolmusz 2000, 2002), Ramsey graphs (Gopalan 2006b; Grolmusz 2000) and locally decodable codes (Efremenko 2009). These results require low-degree *weak* representations of simple Boolean functions like the Or function but modulo composites.

Definition 1.7. *The polynomial* $P(x_1, \ldots, X_n) \in \mathbb{Z}[X_1, \ldots, X_n]/m\mathbb{Z}$ *weakly represents* $f : \{0,1\}^n \to \{0,1\}$ *over* $\mathbb{Z}/m\mathbb{Z}$ *if* $f(x) \neq f(y) \Rightarrow P(x) \neq P(y)$ *(*$P(x)$ *may take values in* $\mathbb{Z}/m\mathbb{Z}$*).*

Such representations have been well studied in complexity theory (see Barrington *et al.* 1994; Bhatnagar *et al.* 2006 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 $O(\sqrt{n})$ (Barrington *et al.* 1994) versus $\Omega(\log n)$ (Tardos & Barrington 1998) between upper and lower bounds. Better upper bounds would lead to improved constructions of all the above combinatorial objects. In Gopalan (2006b), Gopalan proposes viewing this as a question about the degree of two related functions in distinct characteristics:

Problem 1.8 (Gopalan 2006b). *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 $\mathbb{F}_p$ on n variables, there exist a constant t and a linear combination of the form

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

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 (Bogdanov & Viola 2007; Lovett 2008; Viola 2008).

With this lemma in hand, one would like to proceed as follows: suppose $P(x)$ and $Q(x)$ represent the same function f over $\mathbb{F}_p$ and $\mathbb{F}_q$, and that $P(x)$ has low degree (say a constant). The polynomial $P'(x)$ is tightly related to the Mod_p function, which is known to require high degree in characteristic q. We would like to claim that the degree of $P'(x)$ over $\mathbb{F}_q$ is a small multiple of $\deg(Q)$, which would then imply that $\deg(Q)$ must be large. Implementing this scheme runs into an obstacle: P' is a function that maps $\mathbb{F}_p^n \to \mathbb{F}_p$, further the values a_i are from $\mathbb{F}_p^n$, 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. The special case of $p = 2$ is easier to handle, as since $\{0, 1\} \subset \mathbb{F}_q$ one can mimic operations modulo 2 in characteristic $\mathbb{F}_q$ without a large overhead. we present the case of characteristic 2 separately in Section 4. For $p > 2$, we show that one can still mimic differentiation modulo p in characteristic q without a large blowup in the degree, however the argument is more complicated. We present 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 $\mathbb{Z}$, $\mathbb{Z}/m\mathbb{Z}$ 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 $f : \{0, 1\}^n \to R$:

$$P(x) = \sum_{S \subseteq [n]} c_S \prod_{i \in S} x_i$$

$$\text{where} \quad c_S = \sum_{x \leq x(S)} (-1)^{|S| - \mathrm{wt}(x)} f(x)$$

where $x(S)$ denotes the indicator vector of the set S, $x \leq x(S)$ denotes that $x_i \leq x(S)_i$ for every coordinate i and $\mathrm{wt}(x)$ 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 Gopalan (2006a). The multilinear polynomial computing f over $\mathbb{Z}/m\mathbb{Z}$ 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* $\deg_m(f) \leq \deg(f)$ *for all* m. *Similarly if* $m_1 | m$, *then* $\deg_{m_1}(f) \leq \deg_m(f)$.

A consequence of this inequality is that $\deg_m(f) \leq \deg_{m^k}(f)$. 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:*

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

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

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

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

Thus if we know $\deg_p(f)$ for all primes p that divide m, we can use Fact 2.2 and Fact 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 $\mathrm{AC}_0[p]$ circuit, then f can be well approximated by low degree polynomials over $\mathbb{F}_p$. The first is their low-degree approximation lemma for $\mathrm{AC}_0[p]$ circuits.

LEMMA 2.5 (Razborov 1987; Smolensky 1987). *For a prime p, let f be a Boolean function on n variables that is computed by an $\mathrm{AC}_0[p]$ 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*

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

for some absolute constant c.

The second lemma shows that the Mod_p function does not have such an approximation over $\mathbb{F}_q$.

Lemma 2.6 (Razborov 1987; Smolensky 1987). *For every two primes $p \neq q$, there exist constants $c, \epsilon > 0$ depending only on p, q such that for any polynomial $Q(x)$ over $\mathbb{F}_q$ of degree at most $c\sqrt{n}$,*

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

We do not care about exact constants in this paper, unless otherwise specified. Hence, to simplify notation we denote constants by c, where we specify whether these are absolute constants or depending on some other parameters (i.e. ϵ, p, q). In all cases constants do not depend on the number of variables n.

2.1. Proof of Lemma 1.4. For completeness we give the simple proof of Lemma 1.4. The proof follows the Nisan–Szegedy argument, which gives upper and lower bounds on the average sensitivity of the Boolean function in terms of $\deg(f)$. We observe that the lower bound holds in any characteristic (but the upper bound holds only for characteristic 0).

Proof (Proof of Lemma 1.4). Let us define $\mathbb{I}\mathrm{nf}_i(f) = \Pr_{x\in\{0,1\}^n}[f(x) \neq f(x \oplus e_i)]$ where $x \oplus e_i$ denotes x with the i^{th} bit flipped. A simple application of the Schwartz-Zippel lemma shows that

$$\mathbb{I}\mathrm{nf}_i(f) \geq \frac{1}{2^{\deg_p(f)}} \quad \text{hence} \quad \sum_{i \leq n} \mathbb{I}\mathrm{nf}_i(f) \geq \frac{n}{2^{\deg_p(f)}}\,.$$

But by Corollary 1 in Nisan & Szegedy (1992),

$$\sum_{i \leq n} \mathbb{I}\mathrm{nf}_i(f) \leq \deg(f)$$

which gives the required bound. □

3. Degree reduction

A crucial tool in our proofs is the following *Degree reduction lemma* that reduces degree d polynomials in n variables to polynomials with many linear terms. For a polynomial P define the set $L(P)$ to be those variables x_i appearing as linear terms in P but not in any of its higher degree monomials.

Lemma 3.1 (Degree Reduction Lemma). *Let $P(x)$ be a polynomial of degree d over $\mathbb{F}_p$, depending on all n variables, such that the individual degree of each variable is at most $p-1$. Then there exist $t \leq p^{\lceil \frac{d-1}{p-1} \rceil}$, $a_1, \dots, a_t \in \mathbb{F}_p^n$, and $\lambda_1, \dots, \lambda_t \in \mathbb{F}_p$ 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. The main idea used is that if $P(x)$ is a homogeneous degree d polynomial, then taking $d-1$ directional derivatives of P along random directions will yield with high probability a polynomial with many linear variables. In the non-homogenous case, we have to choose how many times to differentiate carefully, since for example if the polynomial is $X_1X_2 + X_3 + X_4 \cdots + X_n$, then most of the variables will disappear after differentiating just once. To get a large linear form from this polynomial however, we can simply set $X_1 = X_2 = 0$. Our final degree reduction procedure combines these two strategies, we first differentiate and then set some variables to 0 to get a large linear form.

Finally, for technical reasons, we differentiate multiple times along each direction rather than choosing multiple directions. While this makes the proof of the degree reduction more involved, it allows us to get a better dependence on the degree. Roughly speaking, we can show that $\deg_q(f) \geq \frac{n}{p^{\deg_p(f)}}$, whereas differentiating once along multiple directions would yield bounds of the form $\deg_q(f) \geq \frac{n}{2^{p \deg_p(f)}}$ with our proof technique.

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 the 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 derive a formula for those 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 3.2. *Let* $\ell \leq p - 1$. *Then*

$$\begin{aligned} &\mu(\ell, c) = 0 \quad \textit{for} \quad c \in \{0, \ldots, \ell - 1\}\,, \\ &\mu(\ell, \ell) \not\equiv 0 \mod p\,. \end{aligned}$$

PROOF. We prove the first identity by induction on c. The case $c = 0$ is elementary. To prove it for $c \geq 1$, we consider the following identity over $\mathbb{Z}$

$$(X - 1)^\ell = \sum_{0 \leq j \leq \ell} (-1)^{\ell-j} \binom{\ell}{j} X^j\,. \tag{3.3}$$

Differentiating both sides $c \leq \ell - 1$ times and then setting $X = 1$ gives

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

where the $\lambda(i)$-s are some integers. Using the induction hypothesis for $i \leq c-1$ gives $\mu(\ell, c) = 0$. To prove $\mu(\ell, \ell) \not\equiv 0 \mod p$ we differentiate Equation (3.3) ℓ times to get

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

Since we assume that $\ell \leq p - 1$ it follows that $\mu(\ell, \ell) = \ell! \not\equiv 0 \bmod p$. □

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 say $e \leq d$ if $e_i \leq d_i$ for all i, and use the notation $\binom{d}{e} = \prod_i \binom{d_i}{e_i}$. We have

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

where we use $\mu(\ell, |e|) = 0$ for $|e| \leq \ell - 1$. 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 $\prod_{l \in [n]} \binom{d_l}{e_l^{(1)}, \ldots, e_l^{(k)}}$. We have

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

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

$$\begin{aligned}
& P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}(x) = \sum_d c_d x^d_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})} \\
&= \sum_d c_d \sum_{|e^{(1)}| \geq \ell^{(1)}, \ldots, |e^{(k)}| \geq \ell^{(k)}} \binom{d}{e^{(1)}, \ldots, e^{(k)}} x^{d-(\sum_j e^{(j)})} \cdot \prod_{j=1}^{k} \mu(\ell^{(j)}, |e^{(j)}|) (y^{(j)})^{e^{(j)}}
\end{aligned}$$

(3.4)

$$= \sum_f x^f \left(\sum_{|e^{(1)}| \geq \ell^{(1)}, \ldots, |e^{(k)}| \geq \ell^{(k)}} c_{f+\sum_j e^{(j)}} \binom{f + \sum_j e^{(j)}}{e^{(1)}, \ldots, e^{(k)}} \cdot \prod_{j=1}^{k} \mu(\ell^{(j)}, |e^{(j)}|)(y^{(j)})^{e^{(j)}} \right)$$

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 $\min(\sum_j \ell^{(j)}, \deg_i(P))$ (note that the degree cannot drop below zero):

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

Lemma 3.5. *Let*

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

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

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. Consider the vector f where $f_i = 1$ and $f_j = 0$ for all $j \neq i$. By Equation (3.4), the coefficient of x^f in $P_{(y^{(1)},\ell^{(1)}),\ldots,(y^{(k)},\ell^{(k)})}(x)$ is given by

(3.6)

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

We shall now find $e^{(1)}, \ldots, e^{(k)}$ so that the following conditions hold:

$$c_{f+\sum_j e^{(j)}} \neq 0, \quad \binom{f + \sum_j e^{(j)}}{e^{(1)}, \ldots, e^{(k)}} \neq 0 \tag{3.7}$$

$$|e^{(1)}| = \cdots = |e^{(k-1)}| = p - 1, \quad |e^{(k)}| = \ell\,. \tag{3.8}$$

Indeed, Equation (3.8) ensures that $\mu(\ell_j, |e^{(j)}|) \neq 0$. By Equation (3.7) 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. Notice that distinct solutions contribute distinct monomials to the right hand side of (3.6). Hence, the claim will follow if we show that there is at least one choice of $e^{(1)}, \dots, e^{(k)}$ satisfying Equations (3.7), (3.8).

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

$$|e^{(1)}| = \cdots = |e^{(k-1)}| = p - 1, |e^{(k)}| = \ell$$

and

$$\sum_j (e^{(j)})_l + f_l = d_l \quad \forall\, l \in [n] .$$

Note that

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

As

$$\sum_j (e^{(j)})_l \leq f_l + \sum_j (e^{(j)})_l = d_l \leq p - 1 ,$$

each binomial coefficient in the product is non-zero mod p. This gives a solution satisfying both Equations (3.7) and (3.8). □

Let $\delta_p(d)$ denote the minimum probability that a nonzero degree d polynomial over $\mathbb{F}_p$ evaluates to zero on a random input. It is well-known (see e.g. MacWilliams & Sloane 1977) 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 3.9. *Let $P(x) \in \mathbb{F}_p[x]$ be a degree d polynomial that depends on all n variables. Then there exist $k \leq \lceil \frac{d-1}{p-1} \rceil$, directions $y^{(1)}, \dots, y^{(k)} \in \mathbb{F}_p^n$ and integers $\ell^{(1)}, \dots, \ell^{(k)} \leq p - 1$ such that*

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

Proof. There 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 trivially holds, so assume

$d' > 1$. Let $d'-1 = (k-1)(p-1)+\ell$ for $\ell \le p-2$ and set $\ell^{(1)} = \cdots = \ell^{(k-1)} = p-1, \ell^{(k)} = \ell$. Lemma 3.5 implies that, 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 $d'-1 \le d-1$ in $y^{(1)}, \ldots, y^{(k)}$. Thus, there exists a setting for $y_1, \ldots, y_k$ where at least

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

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

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

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

4. The case of characteristic 2

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

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

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

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

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

$$\deg_q(g) \le k \deg_q(f).$$

Proof. For any $a \in \mathbb{F}_2^n$, consider $f_a(x) = f(x \oplus a)$. Clearly, $g(x) = \oplus_{i=1}^k f_{a_i}(x)$. We claim that $\deg_q(f_a) = \deg_q(f)$. Let $Q(x)$ be a polynomial over $\mathbb{F}_q$ 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 $\deg_q(Q_a) = \deg_q(Q)$.

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

We now restate and prove Theorem 1.2 in the $p = 2$ case, showing that any Boolean function with small degree over $\mathbb{F}_2$ must have high degree over $\mathbb{F}_q$ 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 such that $\deg_2(f) = d$. Let $P(x)$ be the degree d polynomial over $\mathbb{F}_2$ computing f. We will prove that the multilinear polynomial $Q(x)$ over $\mathbb{F}_q$ computing f has high degree.

By Lemma 3.1, there exist $a_1, \dots, a_k \in \mathbb{F}_2^n$, for $k \leq 2^d$, such that if $\tilde{P}(x) = \oplus_{i=1}^k P(x \oplus a_i)$, then $|L(\tilde{P})| \geq \frac{n}{d2^d}$. Let us denote the set $L(\tilde{P})$ by S. Let $\tilde{P}_S$ be the restriction of $\tilde{P}$ to the variables in S obtained by fixing the remaining variables to zero. Clearly, $\tilde{P}_S(x)$ 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 $\tilde{Q}$ defined as $\tilde{Q}(x) = H^{\oplus}(Q(x \oplus a_1), \dots, Q(x \oplus a_k))$ satisfies that $\tilde{Q}(x) = \tilde{P}(x)$ for all $x \in \{0,1\}^n$. So if we let $\tilde{Q}_S$ be the restriction of $\tilde{Q}$ to the variables in S, then $\tilde{Q}_S(x) = \tilde{P}_S(x)$ for all $x \in \{0,1\}^n$.

Now, since $\tilde{P}_S$ is the parity function over $|S|$ bits, Claim 4.1 implies that $\deg(\tilde{Q}_S) = |S| \geq \frac{n}{d2^d}$. On the other hand, by Lemma 4.2 we have that $\deg(\tilde{Q}_S) \leq \deg(\tilde{Q}) \leq k \deg_q(f)$. Therefore 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 $\mathbb{F}_q$. We need the following claim, which is proven using the union bound.

Claim 4.4. *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 $a_1, \ldots, a_k \in \mathbb{F}_2^n$. Then*

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

We now restate and prove Theorem 1.5 in the $p=2$ case.

Theorem 4.5 (Theorem 1.5, $p=2$ case). *For a prime $q \neq 2$ let $c, \epsilon > 0$ be given by Lemma 2.6. Let $f : \{0,1\}^n \to \{0,1\}$ be of degree $\deg_2(f) = d$. If $h : \mathbb{F}_q^n \to \mathbb{F}_q$ satisfies*

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

then

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

Proof. Using Lemma 3.1, choose $k \leq 2^d$ and $a_1, \ldots, a_k \in \mathbb{F}_2^n$ so that there exists a set of variables S of size $|S| \geq \frac{n}{d2^d}$ such that the function $\tilde{f}(x) = \oplus_{i=1}^k f(x\oplus a_i)$ is either Parity or its negation when restricted to the variables in S. Similarly, define $\tilde{h}(x) = \oplus_{i=1}^k h(x\oplus a_k)$. By Claim 4.4 we get that

$$\Pr_{x\in\{0,1\}^n}\left[\tilde{f}(x) = \tilde{h}(x)\right] \geq 1 - k2^{-d}\epsilon \geq 1-\epsilon\,.$$

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

$$\Pr_{x\in\{0,1\}^S}\left[\tilde{f}_{S,b}(x) = \tilde{h}_{S,b}(x)\right] \geq 1-\epsilon\,.$$

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

$$\mathbf{E}_{b\in\{0,1\}^{[n]\setminus S}}\left[\Pr_{x\in\{0,1\}^S}\left[\tilde{f}_{S,b}(x) = \tilde{h}_{S,b}(x)\right]\right] = \Pr_{x\in\{0,1\}^n}\left[\tilde{f}(x) = \tilde{h}(x)\right] \geq 1-\epsilon\,.$$

We also have $\deg_q(\tilde{h}_{S,b}) \leq \deg_q(\tilde{h}) \leq 2^d \deg_q(h)$, where the last inequality uses Lemma 4.2. Now, $\tilde{f}_{S,b}(x)$ is either Parity or its negation (assume w.l.o.g the former) over $|S|$ variables. Since $\tilde{h}_{S,b}$ approximates Parity over $|S|$ variables with probability at least $1-\epsilon$, Lemma 2.6 implies $\deg_q(\tilde{h}_{S,b}) \geq c\sqrt{|S|}$. Thus

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

which proves the theorem. □

Combining Theorem 4.5 with the Razborov–Smolensky bound, we conclude that any $\mathrm{AC}_0[q]$ circuit that computes a low $\mathbb{F}_2$-degree Boolean function on n variables must be of exponential size.

THEOREM 4.6 (Theorem 1.6, $p = 2$ case). *For any prime $q \neq 2$, there exist constants c_1, c_2 so that any $\mathrm{AC}_0[q]$ circuit of depth t computing a function $f : \{0,1\}^n \to \{0,1\}$ on n variables with $\deg_2(f) = d$ requires size $c_1 2^{-d} \exp((c_2 \frac{n}{d8^d})^{\frac{1}{2t}})$.*

PROOF. Assume there is an $\mathrm{AC}_0[q]$ circuit of size s and depth t computing f. Let ϵ be the constant in Lemma 2.6. Applying Lemma 2.5 with $\delta = 2^{-d}\epsilon$, there is some absolute constant c' and an $\mathbb{F}_q$ polynomial Q of degree $\deg(Q) \leq (c' \log \frac{s}{2^{-d}\epsilon})^t$ such that

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

By Theorem 4.5 we get that $\deg(Q) \geq c\sqrt{\frac{n}{d8^d}}$ for some constant c. Hence,

$$s \geq c_1 2^{-d} \exp\left(\left(c_2 \frac{n}{d8^d}\right)^{\frac{1}{2t}}\right),$$

for absolute constants c_1, c_2. □

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.

In this section we work with polynomials that represent a Boolean function over different characteristics. Suppose f is a Boolean function with low degree over $\mathbb{F}_p$. Our goal is to show that some suitable derivative of f is a linear function. We will then try to relate the degree of this derivative over $\mathbb{F}_q$ to $\deg_q(f)$. This scheme becomes harder to implement, since in differentiating a polynomial over $\mathbb{F}_p^n$, we need to take linear combinations of various points in $\mathbb{F}_p^n$. There is no natural way to associate $\mathbb{F}_p^n$ with a subset of $\mathbb{F}_q^n$ for $p > 2$. To overcome this difficulty, we define a suitable embedding of $\mathbb{F}_p^n$ to $\mathbb{F}_q^n$. While the proof is now technically harder, the basic idea stays the same.

Let $f(x)$ be a Boolean function. We start by defining a polynomial extending f to a function $F : \mathbb{F}_p^n \to \{0,1\}$. Given a vector $x \in \mathbb{F}_p^n$, 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 $F : \mathbb{F}_p^n \to \{0,1\}$ by $F(x) = f(x^{p-1})$. $F(x)$ can be expressed as a polynomial of degree $(p-1)\deg_p(f)$ by considering the multilinear representation of f over $\mathbb{F}_p$ and replacing each variable x_i with x_i^{p-1}; henceforth we shall think of F as this polynomial. Our goal will be to show that if f has low degree over $\mathbb{F}_q$, 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 $\mathbb{F}_p^n$, we need to define the notion of computing functions on $\mathbb{F}_p^n$ by polynomials over $\mathbb{F}_q$. 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 $\mathbb{F}_p^n$ with a subset of $\mathbb{F}_q^{nb}$ by identifying $x = (x_1,\ldots,x_n) \in \mathbb{F}_p^n$ with $(x_{1,1},\ldots,x_{1,b},\ldots,x_{n,1},\ldots,x_{n,b}) \in \mathbb{F}_q^{nb}$, where the value of x_i determines the values of $(x_{i,1},\ldots,x_{i,b})$. Notice that in fact we map $\mathbb{F}_p^n$ into $\{0,1\}^{nb} \subset \mathbb{F}_q^{nb}$. Given $x \in \mathbb{F}_p^n$, we use $\bar{x} \in \{0,1\}^{nb}$ to denote the vector in $\{0,1\}^{nb} \subset \mathbb{F}_q^{nb}$ that represents it. We use $\bar{x_i}$ to denote the vector $(x_{i,1},\ldots,x_{i,b})$ representing x_i. We say that a polynomial $G(x) \in \mathbb{F}_q[x_{1,1},\ldots,x_{n,b}]$ computes $F : \mathbb{F}_p^n \to \{0,1\}$ if $F(x) = G(\bar{x})$ for every $x \in \mathbb{F}_p^n$. We now show that if f has low degree in $\mathbb{F}_q$, then $F(x +_p a)$ can also be computed by a low degree polynomial over $\mathbb{F}_q$.

LEMMA 5.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. Let $F(x)$ be a polynomial over $\mathbb{F}_p$ defined by $F(x) = f(x^{p-1})$. Then, for every $a \in \mathbb{F}_p^n$ there is a polynomial $G_a(x) \in \mathbb{F}_q[x_{1,1},\ldots,x_{n,b}]$ over $\mathbb{F}_q$ of degree at most $b \cdot \deg_q(f)$ computing $F(x +_p a)$.*

PROOF. For $a = (a_1,\ldots,a_n) \in \mathbb{F}_p^n$ and $i \in [n]$ let $A_i(\bar{x_i}) \in \mathbb{F}_q[\bar{x_i}]$ be such that $\deg(A_i) \le b$ and

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

Recall that $\bar{x_i}$ is a 0/1 vector of length b, therefore we can define A_i to be a multilinear polynomial by only considering its values on $\{0,1\}^b$. When the input to A_i is not a vector of the form $\bar{x_i}$ we allow it to output an arbitrary value in $\mathbb{F}_q$. As A_i is multilinear its degree is clearly at most b. By definition it follows that $(A_1(\bar{x_1}),\ldots,A_n(\bar{x_n})) = (x +_p a)^{p-1}$. Let $g : \mathbb{F}_q^n \to \mathbb{F}_q$ be a polynomial of degree $\deg_q(f)$ representing f over $\mathbb{F}_q$. Define the polynomial $G_a(\bar{x}) : \mathbb{F}_q^{bn} \to \mathbb{F}_q$ as

$$G_a(\bar{x}) = g\big(A_1(\bar{x_1}),\ldots,A_n(\bar{x_n})\big)\,.$$

We have:

$$G_a(\bar{x}) = g\big(A_1(\bar{x_1}),\ldots,A_n(\bar{x_n})\big) = g\big((x+_p a)^{p-1}\big) = f\big((x+_p a)^{p-1}\big) = F(x+_p a)$$

as required, and $\deg(G_a) \le b\deg(g) = b\deg_q(f)$. □

As in the proof of Lemma 4.2 we shall need to compute Boolean predicates, on expressions of the form $F(x +_p a_1) +_p \cdots +_p F(x +_p a_k)$, by low degree polynomials over $\mathbb{F}_q$.

COROLLARY 5.2. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $F(x)$ be a polynomial over $\mathbb{F}_p$ defined by $F(x) = f(x^{p-1})$. Let $a_1, \ldots, a_k \in \mathbb{F}_p^n$, $\lambda_1, \ldots, \lambda_n \in \mathbb{F}_p$ and $t : \mathbb{F}_p \to \{0,1\}$ be any Boolean valued predicate on $\mathbb{F}_p$. Define the function $T : \mathbb{F}_p^n \to \{0,1\}$ as*

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

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

PROOF. By Lemma 5.1, each function $F(x +_p a_i) = f((x +_p a_i)^{p-1})$ can be computed by a polynomial $G_i(\bar{x})$ over $\mathbb{F}_q$ 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 $\mathbb{F}_q$ computing the function $t(\lambda_1 z_1 +_p \cdots +_p \lambda_k z_k) : \{0,1\}^k \to \{0,1\}$. Thus, T can be computed by a polynomial over $\mathbb{F}_q$ 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 $a_1, \ldots, a_k \in \mathbb{F}_p^n$ 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 $s = |S|$. Let G_S be the restriction of G to the variables in S (by setting the other variables to zero). We get that for some $\alpha_1, \ldots, \alpha_s \in \mathbb{F}_p \setminus \{0\}$ and $\alpha_0 \in \mathbb{F}_p$,

$$G_S(x) = \sum_{i=1}^{s} \alpha_i x_i + \alpha_0 .$$

Let ω be a p^{th} root of unity in the appropriate extension field $\mathbb{F} = \mathbb{F}_{q^r}$ of $\mathbb{F}_q$. We consider the function $h : \{0,1\}^s \to \mathbb{F}$, which, by abuse of notations, is given by $h(x) = \omega^{\sum_{1\leq i\leq s} \alpha_i x_i +_p \alpha_0}$. Indeed, we think of the expression $\sum_{1\leq i\leq s} \alpha_i x_i +_p \alpha_0$ as taking values in $\{0, 1, \ldots, p-1\}$ and then raise ω to the appropriate power.

The unique multilinear polynomial $H(x)$ over $\mathbb{F}$ computing h on $\{0,1\}^s$ has degree $\deg_{\mathbb{F}}(H) = s \geq \frac{n}{dp^d}$ and is given by

$$H(x) = \omega^{\alpha_0} \prod_{i=1}^{s} \left(1 + (\omega^{\alpha_i} - 1)x_i\right).$$

We now upper-bound $\deg(H)$ in terms of $\deg_q(f)$. First, for $i \in \{0, \ldots, p-1\}$ let $t_i : \mathbb{F}_p \to \{0,1\}$ be the predicate indicating whether $x \equiv i \bmod p$. Consider the polynomial $T_i : \mathbb{F}_p^n \to \{0,1\}$ defined by $T_i = t_i(G_S(x))$. Since $G_S(x)$ is obtained by setting some of the variables in $\sum_i \lambda_i F(x +_p a_i)$ to zero, Corollary 5.2 gives $\deg_q(T_i) = \deg_q(t_i(G_S(x))) \leq kb \deg_q(f)$. Notice that as $H(x)$ is unique, it also equal to the multlinearization of the polynomial

$$\tilde{H}(x) = \sum_{i=0}^{p-1} \omega^i T_i(x).$$

It follows that

$$s = \deg_{\mathbb{F}}(H) \leq \max_i \deg_q \left(T_i(x)\right) = \max_i \deg_q \Big(t_i\big(G_S(x)\big)\Big) \leq kb \deg_q(f).$$

Therefore,

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

We use Theorem 1.2 to prove Corollary 1.3.

Proof of Corollary 1.3. Let p be the smallest prime divisor of m and let $q \neq p$ be another prime divisor. Note that by Fact 2.3, we have $\deg_m(f) \geq \max(\deg_p(f), \deg_q(f))$ so it suffices to show that one of $\deg_p(f)$ or $\deg_q(f)$ exceeds the claimed bound.

So assume that $\deg_p(f) \leq \frac{1}{2}\log_p n - \log_p \log_p n - \frac{1}{2}\log_p \lceil \log_2 p \rceil$. By Theorem 1.2, we get

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

where the last inequality is a simple calculation. This proves the desired bound. □

Next we prove Theorem 1.5 showing that functions with low degree over $\mathbb{F}_p$ are hard to approximate over $\mathbb{F}_q$. First we state the theorem precisely.

Theorem 5.3 (Theorem 1.5 for general p). *For any primes $p \neq q$ there exist constants $c, \epsilon > 0$ depending only on p, q such that the following holds. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function depending on all n variables with $\deg_p(f) = d$. Let $h : \mathbb{F}_q^n \to \{0,1\}$ be any function satisfying*

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

Then

$$\deg_q(h) \geq c\sqrt{\frac{n}{dp^{3d}}}\,.$$

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 $F_v : \mathbb{F}_p^n \to \{0,1\}$ 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 $\mathbb{F}_p$ 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 $v_i = 0$. As this argument appeared several times before we omit the details.

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

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

Then there exists $v \in \{0,1\}^n$ such that if we define $F_v(x) = f(x^{p-1} \oplus v)$ and $G_v = g(x^{p-1} \oplus v)$ then

$$\Pr_{x \in \mathbb{F}_p^n} \big[F_v(x) = G_v(x)\big] \geq 1 - \epsilon\,.$$

Proof. Consider the following expression over a uniform choice of $v \in \{0,1\}^n$

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

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

We also need the following analogue of Claim 4.4:

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

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

We now prove Theorem 5.3.

Proof of Theorem 1.5 in the case of general p. Let $f(x)$ be a Boolean function of small degree d over $\mathbb{F}_p$. Let $h(x) : \mathbb{F}_q^n \to \{0,1\}$ be such that $\Pr_{x\in\{0,1\}^n}[f(x) = h(x)] \geq 1 - p^{-d}\epsilon$, for a small enough $\epsilon > 0$. We will prove that $\deg_q(h)$ 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 h will be transformed into a function whose degree over $\mathbb{F}_q$ is bounded in terms of $\deg_q(h)$. By Lemma 2.6, it must then follow that $\deg_q(h)$ is large. From this point on we shall 'forget' that h is defined over $\mathbb{F}_q^n$ and only consider its values on $\{0,1\}^n \subset \mathbb{F}_q^n$. In other words, we shall think of h as a Boolean function.

The first step is to extend f, h to functions mapping $\mathbb{F}_p^n$ to $\{0,1\}$. Let $F_v(x) = f(x^{p-1} \oplus v)$ and $H_v(x) = h(x^{p-1} \oplus v)$ be mappings from $\mathbb{F}_p^n$ to $\{0,1\}$. By Claim 5.5, there exists $v \in \{0,1\}^n$ such that

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

In addition, the degree of F_v over $\mathbb{F}_p$ is at most $(p-1)d$. The next step is to apply the degree reduction lemma to F_v. By Lemma 3.1, there is some k where

$$k \leq p^{\lceil \frac{\deg(F_v)-1}{p-1} \rceil} \leq p^d$$

vectors $a_1,\dots,a_k \in \mathbb{F}_p^n$ and $\lambda_1,\dots,\lambda_n \in \mathbb{F}_p$, such that for $G_f(x) = \sum_{i\leq k} \lambda_i F_v(x +_p a_i)$ (the sum is addition modulo p) it holds that the set $S = L(G_f)$ has size $s \geq \frac{n}{dp^d}$. Let $G_h : \mathbb{F}_p^n \to \mathbb{F}_p$ be defined as

$$G_h(x) = \sum_{i\leq k} \lambda_i H_v(x +_p a_i)\,. \tag{5.7}$$

Claim 5.6 implies that

$$\Pr_{x\in\mathbb{F}_p^n}\big[G_f(x)=G_h(x)\big]\geq 1-kp^{-d}\epsilon\geq 1-\epsilon\,.$$

As in the proof of Theorem 4.5, there exists an assignment $u\in\mathbb{F}_p^{[n]\setminus S}$ to the variables outside S so that the agreement between G_f and G_h is at least as large. To ease notation, we denote these restrictions also as $G_f(x)$ and $G_h(x)$ (instead of $G_{f\,S,u}(x)$ and $G_{hS,u}(x)$). Note that $G_f(x)=\sum_{i\leq k}\alpha_i x_i+_p\alpha_0$ where for $1\leq i\leq s$ $\alpha_i\in\mathbb{F}_p\setminus\{0\}$, $\alpha_0\in\mathbb{F}_p$ and the summation is modulo p. By replacing each x_i in G_f and G_h by $\alpha_i^{-1}x_i$, we get new functions $G'_f, G'_h:\mathbb{F}_p^s\to\mathbb{F}_p$ such that $G'_f(x)=\sum_i x_i+_p\alpha_0$ and

$$\Pr_{x\in\mathbb{F}_p^s}\left[G'_h(x)=\sum_i x_i+_p\alpha_0\right]=\Pr_{x\in\mathbb{F}_p^s}\big[G'_h(x)=G'_f(x)\big]\geq 1-\epsilon\,.$$

The final step is to convert G'_h to a Boolean function approximating the Mod_p function on s variables. Towards this, for each $w\in\mathbb{F}_p^s$, we define $h_w:\{0,1\}^s\to\mathbb{F}_p$ by $h_w(y)=G'_h(y+_p w)$. Note that since $y+_p w$ is distributed uniformly at random over $\mathbb{F}_p^s$ we have that

$$\Pr_{w\in\mathbb{F}_p^s}\left[\Pr_{y\in\{0,1\}^s}\left[h_w(y)=\sum_i y_i+_p\sum_i w_i+_p\alpha_0\right]\right]$$
$$=\Pr_{x\in\mathbb{F}_p^s}\left[G'_h(x)=\sum_i x_i+_p\alpha_0\right]\geq 1-\epsilon\,.$$

Thus there exists w so that

$$\Pr_{y\in\{0,1\}^s}\left[h_w(y)=\sum_{i\leq s} y_i+_p\alpha\right]\geq 1-\epsilon$$
$$\text{where}\quad \alpha=\alpha_0+_p\sum_i w_i\in\mathbb{F}_p\,.$$

Define $t:\mathbb{F}_p\to\{0,1\}$ by $t(z)=1$ iff $z\equiv\alpha \bmod p$ and $t(z)=0$ otherwise. Finally, let $\tilde{h}(y)=t(h_w(y))$. Notice that $t(\sum_{i\leq s}y_i+_p\alpha)=1$ iff $\sum_{i\leq s}y_i\equiv 0 \bmod p$. In other words, $t(\sum_{i\leq s}y_i+_p\alpha)=\mathsf{Mod}_p(y)$. We thus have

$$\Pr_{y\in\{0,1\}^s}\big[\tilde{h}(y)=\mathsf{Mod}_p(y)\big]\geq\Pr_{y\in\{0,1\}^s}\left[h_w(y)=\sum_{i\leq s}y_i+_p\alpha\right]\geq 1-\epsilon\,.$$

Set $\epsilon > 0$ to be the constant guaranteed by Lemma 2.5. By Lemma 2.5, there exist a constant $c' > 0$ (where both c', ϵ depend only on p, q) such that $\deg_q(\tilde{h}) \geq c'\sqrt{s}$. Our goal now is to relate $\deg_q(h)$ to $\deg_q(\tilde{h})$. We make the following observations:

1. We have $h_w(y) = G'_h(y +_p w)$.

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

3. By Equation (5.7), $G_h(x)$ is a linear combination over $\mathbb{F}_p$ of values of the form $H_v(x +_p a_i)$.

4. Each $H_v(x +_p a_i)$ can be computed by a polynomial $Q_i(\bar{x})$ over $\mathbb{F}_q$ of degree at most $\lceil \log_q p \rceil \cdot \deg_q(h)$ by an argument similar to Lemma 5.1.

Thus, we can write $\tilde{h}(y)$ as some predicate $t' : \{0,1\}^k \to \{0,1\}$ applied to a tuple of polynomial $Q_1, \ldots, Q_k$ with $\deg_q(Q_i) \leq \lceil \log_q p \rceil \deg_q(h)$, and hence $\deg_q(\tilde{h}) \leq kb \deg_q(h)$. We conclude that

$$\deg_q(h) \geq \frac{c'\sqrt{s}}{k\lceil \log_q p \rceil} = \frac{c'}{\lceil \log_q p \rceil}\sqrt{\frac{n}{dp^{3d}}}\,.$$

Hence we proved the theorem with the constant $c = \frac{c'}{\lceil \log_q p \rceil}$. □

As a corollary we obtain a lower bound for the size of $AC_0[q]$ circuits computing functions with low degree over $\mathbb{F}_p$.

Theorem 5.8 (Theorem 1.6, restated). *Let p, q be distinct primes. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function depending on all n variables with $\deg_p(f) = d$. Then any $\mathrm{AC}_0[q]$ circuit of depth t computing f requires size at least*

$$c_1 p^{-d} \exp\left(c_2\left(c_3\frac{n}{dp^{3d}}\right)^{\frac{1}{2t}}\right),$$

where c_1, c_2, c_3 are constants depending only on p, q. In particular, for $d = o(\log_p n)$, the lower bound is $\exp(n^{1/2t - o(1)})$.

Proof. Assume there is an $\mathrm{AC}_0[q]$ circuit of size s and depth t computing f. Let ϵ be the constant in Lemma 2.6. Applying Lemma 2.5 with $\delta = p^{-d}\epsilon$ we get that there is some absolute constant c' and an $\mathbb{F}_q$ polynomial $Q : \mathbb{F}_q^n \to \{0,1\}$ of degree $\deg(Q) \leq (c'p \log \frac{s}{p^{-d}\epsilon})^t$ such that $\Pr_{x \in \{0,1\}^n}[Q(x) = f(x)] \geq 1 - p^{-d}\epsilon$.

By Theorem 5.3 $\deg(Q) \geq c''\sqrt{\frac{n}{dp^{3d}}}$ for some constant c'' depending only on p, q. Hence, for $c_1 = \epsilon, c_2 = c'p, c_3 = c''$ we get that

$$s \geq c_1 p^{-d} \exp\left(c_2 \left(c_3 \frac{n}{dp^{3d}}\right)^{\frac{1}{2t}}\right),$$

as claimed. □

6. Open problems

Our work raises some natural questions regarding the relations between $\deg_m(f)$ for various characteristics, some of which we list below:

1. For any integer m, we have $\deg(f) \geq \deg_m(f)$. What is the largest separation possible between these quantities when m is not a prime power? For such m, is $\deg(f)$ polynomial in $\deg_m(f)$? We can restate these questions as follows: Can $\deg(f)$ be bounded as a function of $\deg_p(f)$ and $\deg_q(f)$ for distinct primes p and q?

   Note that the gap between $\deg(f)$ and $\deg_m(f)$ can be unbounded when m is a prime-power. If m is not a prime power, Corollary 1.3 gives an analog of the $\Omega(\log n)$ Nisan–Szegedy lower bound for composite moduli. Thus trivially, $\deg(f)$ is at most exponential in $\deg_m(f)$.

2. The following question was posed by Troy Lee: Given a set S of vectors in $\{0,1\}^n$, define $\mathrm{Rank}_p(S)$ to be the rank of the set S over $\mathbb{F}_p$ and $\mathrm{Rank}(S)$ to be the rank over $\mathbb{R}$. Are there non-trivial relations between these ranks? For example, assume that both $\mathrm{Rank}_2(S)$ and $\mathrm{Rank}_3(S)$ are small, say $\mathrm{poly}(\log n)$. What can be said about $\mathrm{Rank}(S)$? Note that if we consider only $\mathrm{Rank}_2(S)$ then the Hadamard matrix is an example of a full rank matrix over $\mathbb{R}$ that has rank $\log n$ over $\mathbb{F}_2$.

References

J. Aspnes, R. Beigel, M. L. Furst & S. Rudich (1994). The expressive power of voting polynomials. *Combinatorica* **14**(2), 1–14.

D. A. Barrington, R. Beigel & S. Rudich (1994). Representing boolean functions as polynomials modulo composite numbers. *Computational Complexity* **4**, 367–382.

R. Beigel (1993). The Polynomial Method in Circuit Complexity. *Structures in Complexity Theory: 8th Annual Conference* 82–95.

R. Beigel, N. Reingold & D. A. Spielman (1991). The perceptron strikes back. In *Proceedings of the Sixth Conference on Structure in Complexity Theory*, 286–291.

N. Bhatnagar, P. Gopalan & R. J. Lipton (2006). Symmetric polynomials over $\mathbb{Z}_m$ and simultaneous communication protocols. *Journal of Computer and System Sciences* **72**, 252–285.

A. Bogdanov & E. Viola (2007). Pseudorandom bits for polynomials. In 48th *Annual Symposium on Foundations of Computer Science (FOCS'07)*, 41–51. IEEE.

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

K. Efremenko (2009). 3-Query locally decodable codes of exponential codes. In *Accepted to the* 41st *Annual Symposium on the Theory of Computing (STOC'09)*. ACM.

P. Gopalan (2006a). *Computing with Polynomials over Composites*. Ph.D. thesis, Georgia Institute of Technology.

P. Gopalan (2006b). Constructing Ramsey Graphs from Boolean function representations. In *Proceedings of the* 21st *IEEE Conference on Computational Complexity (CCC'06)*.

V. Grolmusz (2000). Superpolynomial Size Set-systems with Restricted Intersections mod 6 and Explicit Ramsey Graphs. *Combinatorica* **20**(1), 71–86.

V. Grolmusz (2002). Constructing set systems with prescribed intersection sizes. *Journal of Algorithms* **44**(2), 321–337.

J. C. Jackson, A. R. Klivans & R. A. Servedio (2002). Learnability beyond AC^0. In *Proceedings of the 34th ACM Symposium on Theory of Computing*.

A. T. Kalai, A. R. Klivans, Y. Mansour & R. A. Servedio (2005). Agnostically Learning Halfspaces. In *Proc.* 46th *IEEE Symp. on Foundations of Computer Science (FOCS'05)*.

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

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

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

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

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

F. J. MacWilliams & N. J. A. Sloane (1977). *The Theory of Error-Correcting Codes*. North-Holland.

M. Minsky & S. Papert (1968). *Perceptrons: an Introduction to Computational Geometry*. MIT Press.

E. Mossel, R. O'Donnell & R. Servedio (2003). Learning Juntas. In *Proceedings of the* 35^{th} *Annual ACM Symposium on the Theory of Computing (STOC'03)*. URL `citeseer.ist.psu.edu/article/mossel03learning.html`.

S. Muroga (1971). *Threshold logic and its applications*. Wiley-Interscience, New York.

N. Nisan & M. Szegedy (1992). 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)*, 462–467. URL `citeseer.ist.psu.edu/nisan92degree.html`.

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

A. A. Razborov (1987). 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.

R. Smolensky (1987). 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)*, 77–82.

G. Tardos & D. A. Mix Barrington (1998). A Lower Bound On The Mod 6 Degree Of The OR Function. *Computational Complexity* **7**, 99–108.

E. Viola (2008). 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).*

Manuscript received 17 September 2009

Parikshit Gopalan
Microsoft Research – Silicon Valley
1065 La Avenida
Mountainview CA 94043, USA
parik@microsoft.com

Shachar Lovett
Faculty of Mathematics and Computer Science
The Weizmann Institute of Science
POB 26, Rehovot 76100, Israel
shachar.lovett@weizmann.ac.il

Amir Shpilka
Department of Computer Science
Technion – Israel Institute of Technology
Haifa 32000, Israel
shpilka@cs.technion.ac.il