COMBINATORICA
Bolyai Society – Springer-Verlag

Combinatorica 1–19

ON THE MINIMAL FOURIER DEGREE OF SYMMETRIC BOOLEAN FUNCTIONS*

AMIR SHPILKA†, AVISHAY TAL‡

Received Jun 19, 2012

In this paper we give a new upper bound on the minimal degree of a nonzero Fourier coefficient in any non-linear symmetric Boolean function. Specifically, we prove that for every non-linear and symmetric $f\colon \{0,1\}^k \to \{0,1\}$ there exists a set $\emptyset \neq S \subset [k]$ such that $|S| = O(\Gamma(k) + \sqrt{k})$, and $\hat{f}(S) \neq 0$, where $\Gamma(m) \leq m^{0.525}$ is the largest gap between consecutive prime numbers in $\{1,\dots,m\}$. As an application we obtain a new analysis of the PAC learning algorithm for symmetric juntas, under the uniform distribution, of Mossel et al. [10].

Our bound on the degree is a significant improvement over the previous result of Kolountzakis et al. [8] who proved that $|S| = O(k/\log k)$.

We also show a connection between lower-bounding the degree of non-constant functions that take values in $\{0,1,2\}$ and the question that we study here.

1. Introduction

One of the most important tools in the analysis of Boolean functions is the Fourier analysis of the function. Roughly, Fourier analysis studies the correlation that the function has with linear functions by applying the discrete Fourier transform. Although the Fourier transform is nothing but a linear transformation on the space of functions, it has found many applications in different areas of theoretical computer science and combinatorics (and

Mathematics Subject Classification (2000): ....fill in, please

* A preliminary version appeared in CCC 2011 [13].

† This research was partially supported by the Israel Science Foundation (grant number 339/10).

‡ This research was partially supported by the Israel Science Foundation (grant number 339/10).

of course other areas of math and physics), a partial list includes learning theory, hardness of approximation, pseudo-randomness, social choice theory, coding theory, cryptography, additive combinatorics and more.

A typical question in Fourier analysis is: given a family of Boolean functions, what can we say about the Fourier spectrum of members in the family. For example, is most of the weight of the Fourier spectrum concentrated on the first few levels? Is the Fourier spectrum spread? Does the function have a nonzero Fourier coefficient at a certain level?

In this paper we consider the family of symmetric Boolean functions and study the following problem: What is the minimal degree such that *any* symmetric Boolean function $f\colon \{0,1\}^k \to \{0,1\}$, which is non-linear over $GF(2)$ (i.e. f is not constant and is not parity nor its negation), has a nonzero Fourier coefficient of (at most) that degree. In other words, what is the minimal size of a set $\emptyset \neq S \subseteq [k]$ such that $\hat{f}(S) \neq 0$.

This problem was first studied (although implicitly) in [10] in the context of giving PAC learning algorithms for Boolean juntas. It was later explicitly discussed in [8], where improved bounds were obtained. A related question was studied in [7]. There the authors studied the question of what is the *maximal* degree such that *any* non-constant symmetric Boolean function $f\colon \{0,1\}^k \to \{0,1\}$ has a nonzero Fourier coefficient of (at least) that degree. Although this question seems the complete opposite of the question that we study here, the two questions are strongly related. Indeed, for $g = f \oplus \text{PARITY}$ we have that $\hat{f}(S) = \hat{g}([k] \setminus S)$. Clearly g is symmetric if and only if f is symmetric. In particular if $\hat{f}(S) = 0$ for all $0 < |S| \leq t$ then it means that g does not have any monomial with degree between $k-t$ and $k-1$. If in addition we know that $\hat{f}(\emptyset) = 0$ then we would get that the degree of g is exactly k minus the minimal size of a set S such that $\hat{f}(S) \neq 0$. Thus, a lower bound on the maximal degree translates to an upper bound on the minimal degree (when $\hat{f}(\emptyset) = 0$). We discuss these results in more detail in Section 1.3.

Besides being a very natural question that continues the investigation of Fourier spectrum of Boolean functions, our work is also motivated by the problem of giving learning algorithms for symmetric juntas.

Learning juntas. One of the most important open problems in learning theory is learning in the presence of irrelevant information. The problem can be described in the following way: we are given as input a set of labelled data points, coming from some high dimensional space and we have to come up with a (small) hypothesis that correctly labels the data points. However, it may be the case that only a small fraction of the data is actually relevant, and so, in order to be able to find such a hypothesis efficiently, we have to discover the relevant variables. This problem appears in many real-life applications. For example, when trying to learn how some genetic attribute depends on the DNA, we expect only a small number of DNA letters to affect this attribute, while the rest are irrelevant.

A nice formulation of the question was proposed by Blum and Langley [4,3]: Let $f\colon \{0,1\}^n \to \{0,1\}$ be an unknown Boolean function depending on $k \ll n$ variables. Henceforth we refer to such a function as a k-junta. We get as input a set of labelled examples $\langle x, f(x)\rangle$ where the data points $x = (x_1, \ldots, x_n)$ are chosen independently and uniformly at random from $\{0,1\}^n$. Our goal is to efficiently identify the k relevant variables and the truth table of the function. It is clear that by going over all $\binom{n}{k}$ possible choices of k variables we can learn f. However, the main question is whether this can be done faster. Specifically, Blum and Langley [3] asked the following, still unsolved, question: "Does there exist a polynomial time algorithm for learning the class of Boolean functions over $\{0,1\}^n$ that have $\log(n)$ relevant features, in the PAC or uniform distribution models?" Note, that for this setting of parameters, this is a sub-problem of the notoriously hard questions of learning polynomial size DNF formulas and decision trees. For more background and applications we refer the reader to [4,3,10]. The results that we obtain in this work improve the analysis of known algorithms for the case where the underlying junta is a symmetric function.

1.1. Our results

Our main result is a new theorem on the degree of the first (non-empty) non-zero Fourier coefficient, of a non-linear symmetric Boolean function f. We shall need the following notation. For an integer m, denote by $\Gamma(m)$ the size of the largest interval inside $\{1, \ldots, m\}$ that does not contain a prime number. In other words,

$$\Gamma(m) = \max\{b - a \mid 1 \le a < b \le m \text{ and there is no prime number in the interval } (a,b]\}.$$

The best bound on Γ was given in [2] where it was shown that $\Gamma(m) \le m^{0.525}$ (for large enough m). We also let $\hat{f}(S)$ be the Fourier coefficient of f at S (see definition in Section 2.1).

Theorem 1.1. *Let $f\colon \{0,1\}^k \to \{0,1\}$ be a non-linear symmetric Boolean function (i.e. f is not constant and is not parity nor its negation). Then, there exists a set $\emptyset \ne S \subset [k]$, of size $|S| = O(\Gamma(k) + \sqrt{k})$, such that $\hat{f}(S) \ne 0$.*

As an application, we give a better analysis for a learning algorithm for the class of symmetric juntas. The symmetric junta learning problem is an interesting sub-case of the general junta learning problem that was first discussed in [10]. In this learning problem, we are guaranteed that the unknown function is symmetric in its k variables. Our result follows by combining the following theorem (implicit in [10,8]) with Theorem 1.1.

Theorem 1.2. *([10,8]) If any non-linear symmetric Boolean k-junta has a non-zero Fourier coefficient of size at most $t(k)$, then the class of symmetric k-juntas can be exactly learned (in the PAC model), from random examples sampled from the uniform distribution, with confidence*[1] $1-\delta$ *in time* $n^{t(k)} \cdot \mathrm{poly}(2^k, n, \log(1/\delta))$.

Corollary 1.3. *The class of symmetric k-juntas over n bits can be , from random examples sampled from the uniform distribution, with confidence $1-\delta$, in time* $n^{O(k^{0.525})} \cdot \mathrm{poly}(2^k, n, \log(1/\delta))$.

Cramér proved that the Riemann hypothesis implies that $\Gamma(m) = O(\sqrt{m}\log m)$ (which is slightly weaker than Legendre's conjecture that $\Gamma(m) = O(\sqrt{m})$) and conjectured that $\Gamma(m) = O(\log^2 m)$ [5]. From our proof technique it follows that if either Cramér's conjecture or Legendre's conjecture is true then Theorem 1.1 may be improved to give a set S of size $O(\sqrt{k})$. This will imply a similar improvement to Corollary 1.3.

Our last result shows a connection between lower-bounding the degree of non-constant functions into $\{0,1,2\}$ and the question of upper bounding the size of the first non-zero Fourier coefficient of a symmetric f.

Theorem 1.4. *If the degree of any non-constant polynomial $H\colon \{0,\dots,k-2\} \to \{-1,0,1\}$ is at least $k-t$, then every non-linear symmetric function $f\colon \{0,1\}^k \to \{0,1\}$ must satisfy $\hat{f}(S) \neq 0$ for some non-empty S of size $|S| < t$.*

1.2. Proof technique

A basic idea that appears in previous work is that if all non-empty Fourier coefficients of f, up to size t, are zero, then no matter how we fix any t variables from f, its bias remains the same. Namely, the probability that f assumes the value 0 is unchanged under any such fixing of at most t variables. This is formally stated in Lemma 3.4. The natural idea now is to consider many different restrictions and to try and combine the information

[1] i.e. probability of outputting an exact hypothesis on uniform random examples

obtained from them to show that the bias cannot remain unchanged, unless f is a linear function.

Denote by $F(i)$ the value that f obtains on inputs that contain exactly i ones and $k-i$ zeros. It follows that $\mathrm{bias}(f)=\frac{1}{2^k}\sum_{i=0}^{k}\binom{k}{i}F(i)$ (see Definition 3.2). If we fix ℓ variables to 1 and $t-\ell$ variables to 0, and the bias is unchanged then we get that $\mathrm{bias}(f)=\frac{1}{2^{k-t}}\sum_{i=0}^{k-t}\binom{k-t}{i}F(i+\ell)$. Assume now that this holds for every $\ell\leq t$, then

$$2^{k-t}\cdot\mathrm{bias}(f)=\sum_{i=0}^{k-t}\binom{k-t}{i}F(i+\ell).$$

Thus, by considering these restrictions we learn that $\{F(i)\}_{i=1,\ldots,k}$ satisfy $t+1$ linear relations. By considering these equations modulo a (well chosen) prime $p\approx k$, and using Lucas' theorem, we are able to simplify them considerably, so that each new equation now contains only two nonzero terms. From these new relations, we deduce that F is either Parity or its negation (cases which are excluded by our assumption that F is non-linear) or that F is fixed on inputs of either very low or very high weight:

$$F(i)=c,\ \text{ for } i\in[0,t]\cup[k-t,k].$$

Considering such relations modulo another prime number implies that F is constant for inputs of weight close to $k/2$. More specifically $F(i)=c$ for $i\in[k/2-\sqrt{k},k/2+\sqrt{k}]$. Then, a simple calculation given in Claim 3.5 and Corollary 3.6 shows that if a function is fixed at such an interval around $k/2$ then it has a nonzero Fourier coefficient of size 1 or 2. This is in contradiction to our assumption that the function has non nonzero Fourier coefficient up to level t.

We note that our proof technique is very similar in nature to that of [8]. There the polynomial $G(z)=F(z+1)-F(z)$ was studied modulo different primes, however the information obtained from those primes was used in a different way than it is used here.

1.3. Related work

In [7], following [11], von zur Gathen and Roche studied the question of giving a lower bound on the degree of non-constant symmetric Boolean functions, when represented as polynomials over the real numbers. In other words, the problem is proving that there is a large set S such that $\hat{f}(S)\neq 0$. They were able to prove that the degree of a symmetric function on k bits is always at least $k-\Gamma(k)$, and conjectured that actually the degree is at least $k-O(1)$. This conjecture is still open. In [6] the related question of providing lower bounds on the degree of symmetric functions from $\{0,1\}^k$ to $\{0,1,\ldots,m\}$ was considered and lower bounds of the form $k-o(k)$ on the degree were proved (when $m<k$). We shall later see the connection between bounding the degree of functions that take values in $\{0,1,2\}$ to proving the existence of a not too large S such that $\hat{f}(S)\neq 0$. We note that the result of [7] actually implies the following corollary. We say that a Boolean function f is balanced if $\Pr_x[f(x)=0]=1/2$. I.e. if f gets the values 0 and 1 equally often. In other words, $\hat{f}(\emptyset)=0$, when f is viewed as a function to $\{-1,1\}$.

Corollary 1.5 ([7]). *Let $f\colon \{0,1\}^k \to \{0,1\}$ be a balanced non-linear symmetric Boolean function. Then, there exists a set $\emptyset\neq S\subset[k]$ of size $|S|\leq \Gamma(k)$ such that $\hat{f}(S)\neq 0$.*

Thus, Theorem 1.1 can be viewed as proving a similar bound for the case of *unbalanced* symmetric functions.

Mossel et al. made the first breakthrough in PAC learning of juntas under the uniform distribution [10]. They gave a learning algorithm whose running time is $n^{\frac{\omega}{\omega+1}k}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$, where ω is the matrix multiplication exponent. Currently the best bound on ω gives $\omega<2.374$ and so their algorithm runs in time (roughly) $n^{0.7k}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$, which is better than the trivial algorithm that runs in time $n^k\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$. This result was improved by Valiant [14] who gave an algorithm whose running time is $n^{\frac{\omega+\epsilon}{4}k}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$, which is better than $n^{0.6k}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$. For the case of symmetric juntas, the algorithm of [10] runs in time an $n^{2k/3}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$. Their analysis for the case of symmetric juntas was significantly improved by Kolountzakis et al. [8] who gave an $n^{O(k/\log k)}\cdot\mathrm{poly}(n,2^k,\log(1/\delta))$ upper bound on the running time of the algorithm for that case. Both results are based on the fact that every non-linear symmetric function f on k variables, has a non-zero Fourier coefficient that is supported on a somewhat small non-empty set S. Namely, on weaker versions of Theorem 1.1.

1.4. Organization

The paper is organized as follows. In Section 2 we give the basic definitions and discuss representations of Boolean functions as polynomials. The proof of Theorem 1.1 is given in Section 3. We prove Theorem 1.4 in Section 4.

2. Preliminaries

We denote $[m] \triangleq \{1,\dots,m\}$. For $x \in \{0,1\}^n$ we denote by $|x|$ the weight of x, i.e., the number of non-zero entries in x. In other words, $|x| = x_1 + \ldots + x_n$. All logarithms in this paper are taken to base 2. We denote by $\binom{n}{\leq r}$ the sum $\sum_{i=0}^{r} \binom{n}{i}$. To ease the reading we will ignore floor and ceiling notations, as it will be obvious that they do not affect the results.

Definition 2.1. A Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ is called a k-junta if it depends on only k of the input bits (usually $k \ll n$). Namely, there exists a function $g\colon \{0,1\}^k \to \{0,1\}$ and k indices $1 \leq i_1 < i_2 < \ldots < i_k \leq n$ such that

$$f(x_1, x_2, \ldots, x_n) = g(x_{i_1}, x_{i_2}, \ldots, x_{i_k})$$

We will be studying integer equations modulo prime numbers and so the following two claims will be useful. The first is the well-known Lucas' theorem.

Theorem 2.2 (Lucas). *Let $a, b \in \mathbb{N} \setminus \{0\}$ and let p be a prime number. Denote by $a = a_0 + a_1 p + a_2 p^2 + \cdots + a_k p^k$ and $b = b_0 + b_1 p + b_2 p^2 + \cdots + b_k p^k$ their base p representations. Then $\binom{a}{b} \equiv_p \prod_{i=0}^{k} \binom{a_i}{b_i}$, where $\equiv_p$ means equality modulo q and $\binom{a_i}{b_i} = 0$ if $a_i < b_i$.*

The second theorem guarantees the existence of a prime number in any large enough interval.

Theorem 2.3 ([2]). *For all large m, the interval $[m - m^{0.525}, m]$ contains prime numbers.*

2.1. Representations of Boolean functions

The basic objects that we study in this paper are symmetric Boolean functions.

Definition 2.4. A function $f\colon \{0,1\}^k \to \{0,1\}$ is symmetric if $f(x) = f(y)$ for all x and y such that $|x| = |y|$.

In other words, a function is symmetric if permuting the coordinates of the input does not change the value of the function.

We shall consider two equivalent ways of representing symmetric Boolean functions. One common and useful representation is the Fourier transform (which applies to non-symmetric functions as well). For this representation it is convenient to think of our function f as mapping $\{-1,1\}^k$ to $\{-1,1\}$, by applying the linear transformation $b\to 1-2b$ (equivalently $b\to(-1)^b$ for $b\in\{0,1\}$) from $\{0,1\}$ to $\{-1,1\}$.

For a subset $S\subseteq[k]$ denote $\chi_S(x)=\prod_{i\in S}x_i$. It is a well-known fact that $\{\chi_S\}_{S\subseteq[k]}$ form an orthonormal basis of the space of functions from $\{-1,1\}^k$ to $\mathbb{R}$ under the inner product $\langle f,g\rangle=\mathbf{E}_{x\in\{-1,1\}^k}[f(x)\cdot g(x)]$, where x is distributed uniformly on $\{-1,1\}^k$. In particular, every Boolean function $f\colon\{-1,1\}^k\to\{-1,1\}$ can be represented as

$$f(x)=\sum_{S\subseteq[k]}\hat{f}(S)\chi_S(x)=\sum_{S\subseteq[k]}\hat{f}(S)\prod_{i\in S}x_i, \tag{1}$$

where

$$\hat{f}(S)=\mathbf{E}_{x\in\{0,1\}^k}[f(x)\cdot\chi_S(x)]. \tag{2}$$

We call $\hat{f}(S)$ the Fourier coefficient of f at S. Note that Equation (1) gives a representation of f as a polynomial over the reals. For example, if we denote $\mathrm{PARITY}_k=\oplus_{i=1}^k x_i$ then, as a polynomial from $\{-1,1\}^k$ to $\{-1,1\}$, we have $\mathrm{PARITY}_k=x_1\cdot x_2\cdots x_k=\chi_{[k]}$. When f is a symmetric polynomial it follows that $\hat{f}(S)=\hat{f}(T)$ whenever $|S|=|T|$. Parseval's identity implies that for $f\colon\{-1,1\}^k\to\{-1,1\}$ it holds that $\sum_{S\subseteq[k]}\hat{f}(S)^2=1$.

Symmetric Boolean functions can also be represented by univariate real polynomials of degree at most k. Indeed, recall that $f(x)$ is actually a function of $|x|=\sum_{i=1}^k x_i$. Hence, there exists a degree $\le k$ polynomial $F\colon\{0,\dots,k\}\to\{0,1\}$ such that $F(|x|)=f(x)$. Similarly to the Fourier representation we shall represent F using a specific basis, $\{1=\binom{x}{0},\binom{x}{1},\binom{x}{2},\dots,\binom{x}{k}\}$. This basis is sometimes called the Newton basis. We can express f as:

$$f(x)=F(|x|)=\sum_{d=0}^{k}\gamma_d\cdot\binom{|x|}{d}. \tag{3}$$

The coefficients γ_d are given in the following lemma (see e.g., problem 36 in [9]).

Lemma 2.5. $\gamma_d=\sum_{i=0}^d(-1)^{d-i}\cdot\binom{d}{i}\cdot F(i)$.

In particular, all γ_d's are integers and γ_d only depends on the first $d+1$ values of F. Another simple fact is that the degree of f as a real polynomial in Equation (1) is the same as the degree of F in Equation (3) (even though the domain is different in the two representations).

It is obvious that f determines F and vice versa and so, in what follows, we shall have both representations in mind and will move freely between them.

We shall denote symmetric functions on the Boolean cube by the letters f, g, h and their corresponding integer polynomials by F, G, H, respectively.

3. Fourier spectrum of symmetric Boolean functions

In this section we prove Theorem 1.1. Our approach is similar to the approach taken by [7,8]. We study the bias of f after restricting some of the variables. From this point on we identify a symmetric function $f\colon \{0,1\}^k \to \{0,1\}$ with its corresponding integer polynomial $F\colon \{0,\dots,k\} \to \{0,1\}$. Recall that $F(i)$ is the value that f obtains on inputs of weight i.

Definition 3.1. Let $F\colon \{0,1,\dots,k\} \to \{0,1\}$ be a symmetric function on k bits. The (m,ℓ)-fixing of F, is a symmetric function on $k-m$ bits $F|_{(m,\ell)}\colon \{0,1,\dots,k-m\} \to \{0,1\}$ defined by

$$F|_{(m,\ell)}(i) \triangleq F(i+\ell).$$

In other words, $f|_{(m,\ell)}$ is the symmetric function obtained by fixing ℓ variables to 1 and $m-\ell$ variables to 0 (again, we identify $f|_{(m,\ell)}$ with $F|_{(m,\ell)}$). We shall study the bias of F under different restrictions.

Definition 3.2. The bias of a function $f\colon \{0,1\}^k \to \{0,1\}$ is defined as $\mathrm{bias}(f) \triangleq \mathbf{E}_{x\in\{0,1\}^k} f(x)$, where x is uniformly distributed.

In other words, the bias is equal to the probability that $f(x)=1$ (when x is picked uniformly at random). In particular, f is unbiased iff $\mathrm{bias}(f)=\frac{1}{2}$. Notice that when f is symmetric then the bias is given by

$$\mathrm{bias}(f) = \frac{1}{2^k}\sum_{i=0}^{k}\binom{k}{i}\cdot F(i).$$

Similarly,

$$\mathrm{bias}(F|_{(m,\ell)}) = \frac{1}{2^{k-m}}\cdot\sum_{i=0}^{k-m}\binom{k-m}{i}\cdot F(i+\ell). \tag{4}$$

The following useful definition and lemma relate the bias of $F|_{(m,\ell)}$ and the Fourier spectrum of f.

Definition 3.3 ([12, 8]). f is called t-null (or t-correlation immune) if for every set $S \subseteq [k]$ such that $1 \le |S| \le t$, it holds that $\hat{f}(S) = 0$.

We remark that the definition above is equivalent to saying that the set $f^{-1}(1)$ is *t-wise-independent* (see definition in [1]), but we do not expand on this connection here.

The notion of correlation immunity was first studied in the cryptography community in the context of security of functions against correlation attacks [12]. We shall use the notation of [8] and refer to such functions as t-null.

Lemma 3.4 ([15, 8]). *The following are equivalent.*

1. *f is t-null.*
2. *For every $0 \le \ell \le m \le t$, $\text{bias}(F|_{(m,\ell)}) = \text{bias}(f)$.*
3. *For every $0 \le \ell \le t$, $\text{bias}(F|_{(t,\ell)}) = \text{bias}(f)$.*

In order to prove that a symmetric f is not t-null, we will look for a (t, ℓ) fixing that changes the bias. Towards this end we shall consider the bias of f modulo different prime numbers. Let $p < k$ be a prime number. If f is $(k-p)$-null then, by Lemma 3.4, there exists c_p such that for all $\ell \le k-p$ it holds that

$$c_p = \text{bias}(F|_{(k-p,\ell)}).$$

In other words, according to Equation (4),

$$2^p \cdot c_p = \sum_{i=0}^{p} \binom{p}{i} \cdot F(i+\ell).$$

Reducing this equation modulo p and using Lucas' theorem (Theorem 2.2) we get that for every $\ell \le k-p$

$$2^p \cdot c_p \equiv_p \sum_{i=0}^{p} \binom{p}{i} \cdot F(i+\ell) \equiv_p F(\ell) + F(p+\ell). \tag{5}$$

Similarly, by considering the case that f is $(k-2p)$-null we get that there exists c_{2p} such that for all $\ell \le k-2p$ it holds that

$$2^{2p} \cdot c_{2p} = \sum_{i=0}^{2p} \binom{2p}{i} \cdot F(i+\ell).$$

As before, reducing modulo p and using Lucas' theorem, we obtain that for every $\ell \le k-2p$

$$2^{2p} \cdot c_{2p} \equiv_p \sum_{i=0}^{2p} \binom{2p}{i} \cdot F(i+\ell) \equiv_p F(\ell) + 2F(p+\ell) + F(2p+\ell). \tag{6}$$

In the next two sections we study the effect of fixing bits on the bias of f and prove Theorem 1.1.

3.1. Fixing 2 bits

In this subsection we present a class of functions for which $\mathrm{bias}(F) \neq \mathrm{bias}(F|_{(2,1)})$. In particular, every function in the class is not 2-null. For $i=1,\ldots,k-1$ the weight of $F(i)$ in $\mathrm{bias}(F)$ is $\frac{1}{2^k}\binom{k}{i}$, whereas the weight of $F(i)$ in $\mathrm{bias}(F|_{(2,1)})$ is $\frac{1}{2^{k-2}}\binom{k-2}{i-1}$. The following is an easy observation.

Claim 3.5.

$$\frac{1}{2^k}\binom{k}{i} \leq \frac{1}{2^{k-2}}\binom{k-2}{i-1} \quad \Leftrightarrow \quad \frac{k-\sqrt{k}}{2} \leq i \leq \frac{k+\sqrt{k}}{2}.$$

Proof. The LHS is equivalent to $k(k-1) \leq 4i(k-i)$. I.e. to $i^2 - ik + k(k-1)/4 \leq 0$. Solving we get the claimed result. ∎

Corollary 3.6. *Let F be a non-constant function $F\colon \{0,1,\ldots,k\} \to \{0,1\}$. If $F(i)=c$ for all $\frac{k-\sqrt{k}}{2} \leq i \leq \frac{k+\sqrt{k}}{2}$, then* $\mathrm{bias}(F) \neq \mathrm{bias}(F|_{(2,1)})$.

Proof. Assume WLOG that $c = 0$. Hence, $F(i) = 1$ only for i such that $i < \frac{k-\sqrt{k}}{2}$ or $\frac{k+\sqrt{k}}{2} < i$, and because F is non-constant there exists some i such that $F(i) = 1$. Thus, the weight of each non-zero $F(i)$ decreases after the fixing, hence the probability that $F = 1$ decreases. Formally,

$$\begin{aligned}
\mathrm{bias}(F) &= \frac{1}{2^k}\sum_{i=0}^{k}\binom{k}{i}F(i) = \frac{1}{2^k}\sum_{i<\frac{k-\sqrt{k}}{2}}\binom{k}{i}F(i) + \frac{1}{2^k}\sum_{\frac{k+\sqrt{k}}{2}<i}\binom{k}{i}F(i) \\
&\geq^{(\dagger)} \frac{1}{2^k}\sum_{1\leq i<\frac{k-\sqrt{k}}{2}}\binom{k}{i}F(i) + \frac{1}{2^k}\sum_{\frac{k+\sqrt{k}}{2}<i\leq k-1}\binom{k}{i}F(i) \\
&\geq^{(*)} \frac{1}{2^{k-2}}\sum_{1\leq i<\frac{k-\sqrt{k}}{2}}\binom{k-2}{i-1}F(i) + \frac{1}{2^{k-2}}\sum_{\frac{k+\sqrt{k}}{2}<i\leq k-1}\binom{k-2}{i-1}F(i) \\
&= \mathrm{bias}(F|_{(2,1)}),
\end{aligned}$$

where inequality $(*)$ follows from Claim 3.5. Observe that there is equality in $(\dagger)$ if and only if $F(0) = F(k) = 0$. Also note that we cannot have that both $(\dagger)$ and $(*)$ are equalities since this would imply that $F(i)=0$ for every i in $\left[0, \frac{k-\sqrt{k}}{2}\right) \cup \left(\frac{k+\sqrt{k}}{2}, k\right]$. This, however, contradicts our assumption that F is not constant. ∎

3.2. Proof of Theorem 1.1

In order to prove Theorem 1.1 we consider restrictions modulo two different primes. The next lemma will be the main tool in the proof of the theorem.

Lemma 3.7. *Let $2 < q \leq k$ be a prime number. Let f be a non-linear symmetric function on k bits which is $(k-q+1)$-null. Then, there exists a constant $c_{q-1} \in \{0,1\}$ such that for every $\ell = 0,\ldots,k-q$*

$$F(\ell) = F(q+\ell) = c_{q-1}.$$

We first show how Theorem 1.1, restated next, follows from the lemma above.

Theorem (Theorem 1.1, restated). *Let $f\colon \{0,1\}^k \to \{0,1\}$ be a non-linear symmetric Boolean function (i.e. f is not constant and is not parity nor its negation). Then, there exists a set $\emptyset \neq S \subset [k]$, of size $|S| = O(\Gamma(k)+\sqrt{k})$, such that $\hat{f}(S) \neq 0$.*

Proof of Theorem 1.1 If f is balanced then the claim follows from Corollary 1.5. Hence, we can assume that f is biased. In addition, assume by contradiction that f is $(2\Gamma(k)+\sqrt{k})$-null. By the definition of Γ, there exist prime numbers p, q such that $\frac{k-\sqrt{k}}{2} - \Gamma(k) \leq p \leq \frac{k-\sqrt{k}}{2}$ and $k-\sqrt{k}-2\Gamma(k) \leq q \leq k-\sqrt{k}-\Gamma(k)$. Since f is $(k-q+1)$-null, Lemma 3.7 implies that there exists a constant $c_{q-1} \in \{0,1\}$ such that

$$F(0) = F(1) = \ldots = F(k-q) = F(q) = F(q+1) = \ldots = F(k) = c_{q-1}. \tag{7}$$

As f is also $(k-2p)$-null, Equation (6) implies that there exists a constant $0 \leq c_{2p} < p$ such that for all $\ell = 0,1,\ldots,k-2p$

$$F(\ell) + 2\cdot F(p+\ell) + F(2p+\ell) \equiv_p c_{2p}.$$

Assuming $4 < p$ (otherwise k is at most some fixed constant and the claim is not interesting), these equations hold over the integers and so we get that for every $\ell = 0,1,\ldots,k-2p$

$$F(\ell) + 2\cdot F(p+\ell) + F(2p+\ell) = c_{2p}. \tag{8}$$

Note that for ℓ such that $\frac{k-\sqrt{k}}{2} \leq p+\ell \leq \frac{k+\sqrt{k}}{2}$, we have

$$\ell \leq \sqrt{k} + \frac{k-\sqrt{k}}{2} - p \leq \sqrt{k} + \Gamma(k) \leq k-q$$

and

$$\ell + 2p \geq \frac{k - \sqrt{k}}{2} + p \geq k - \sqrt{k} - \Gamma(k) \geq q.$$

Combining these observations with Equations (7) and (8) gives

$$2 \cdot F(p+\ell) = c_{2p} - F(\ell) - F(2p+\ell) = c_{2p} - 2c_{q-1}.$$

In other words, F is constant in the interval $\left[\frac{k-\sqrt{k}}{2}, \frac{k+\sqrt{k}}{2}\right]$. By Corollary 3.6 we conclude that f is not 2-null, in contradiction. Therefore, f is not $(2\Gamma(k)+\sqrt{k})$-null, which is what we wanted to prove. ∎

We end this section by proving Lemma 3.7, which concludes the proof of Theorem 1.1.

Proof of Lemma 3.7 Lemma 3.4 implies that since f is $(k-q+1)$-null then for all $\ell = 0, 1, \ldots, k-q+1$ it holds that

$$\sum_{i=0}^{q-1} \binom{q-1}{i} \cdot F(i+\ell) = 2^{q-1} \cdot \text{bias}(F).$$

Consider these equations modulo q. A simple calculation shows that:

$$\binom{q-1}{i} \equiv_q \frac{(-1)\cdot(-2)\cdot \ldots \cdot (-i)}{1 \cdot 2 \cdot \ldots \cdot i} \equiv_q (-1)^i.$$

Therefore, we get that there exists a number $0 \leq c_{q-1} < q$ such that for all $\ell = 0, 1, \ldots, k-q+1$

$$\sum_{i=0}^{q-1} (-1)^i \cdot F(i+\ell) \equiv_q c_{q-1}. \tag{9}$$

Hence, for all $\ell = 0, 1, \ldots, k-q$ it holds that

$$\sum_{i=0}^{q-1} (-1)^i \cdot F(i+\ell) \equiv_q c_{q-1} \equiv_q \sum_{i=0}^{q-1} (-1)^i \cdot F(i+\ell+1).$$

Adding the RHS to the LHS we obtain,

$$\begin{aligned} 2c_{q-1} &\equiv_q F(\ell) + \sum_{i=1}^{q-1} ((-1)^i + (-1)^{i-1}) \cdot F(i+\ell) + (-1)^{q-1} F(q+\ell) \\ &\equiv_q F(\ell) + F(q+\ell). \end{aligned}$$

Thus, $2c_{q-1} \in \{0,1,2\} \mod q$. It follows that c_{q-1} is either $0,1$ or $(q+1)/2$. If $c_{q-1}=0$ or 1 then clearly $F(\ell)=F(q+\ell)=c_{q-1}$ and we are done, so we only need to rule out the case $c_{q-1}=(q+1)/2$. So assume that $c_{q-1}=(q+1)/2$. Equation (9) gives

$$\sum_{i=0}^{q-1}(-1)^i \cdot F(i+\ell) \equiv_q (q+1)/2.$$

In other words,

$$\sum_{i<q:\ i \text{ even}} F(i+\ell) - \sum_{i<q:\ i \text{ odd}} F(i+\ell) \equiv_q (q+1)/2.$$

Therefore it must be the case that either $F(\ell)=F(\ell+2)=\ldots=F(\ell+q-1)=1$ and $F(\ell+1)=\ldots=F(\ell+q-2)=0$, or vice versa. Considering different ℓs shows that F must be parity or its negation, in contradiction to the assumption. ∎

4. On nullity and degree of polynomials taking three values

In this section we prove Theorem 1.4 that gives a connection between the problem of upper bounding the minimal size of a non-zero Fourier coefficient of a symmetric function and the problem of giving a lower bound on the degree of a univariate polynomial $H\colon \{0,\ldots,k\}\to\{0,1,2\}$, that was studied in [6] (in the argument below we consider $H\colon \{0,\ldots,k\}\to\{-1,0,1\}$, but the degrees of H and $H+1$ are of course equal).

Using the observation that $\hat{f}(S)=(f\oplus\text{PARITY})\hat{\ }(S^c)$, where S^c is the complement of S, Mossel et al. [10] concluded that

$$\begin{aligned}(10)\qquad \deg(F) < k-t &\iff \forall S\colon k-t \le |S|, \hat{f}(S)=0\\ &\iff \forall S\colon |S|\le t, (f\oplus\text{PARITY})\hat{\ }(S)=0\\ &\iff f\oplus\text{PARITY}\ \text{ is } t\text{-null and unbiased.}\end{aligned}$$

We first prove a one directional reduction from *any* symmetric t-null function (i.e. even a biased one) to a low degree polynomial that maps $\{0,1,\ldots,k-2\}$ to $\{-1,0,1\}$. To obtain the reduction we need the following lemma that gives a relation between different coefficients, in the Newton basis representation, of a symmetric f such that $f\oplus\text{PARITY}$ is t-null.

Lemma 4.1. *Let $f\colon \{0,1\}^k\to\{0,1\}$ be a symmetric function. If $f\oplus$PARITY is t-null, then, when representing f in Newton's basis, $F(|x|)=\sum_{i=0}^{k}\gamma_i\cdot\binom{|x|}{i}$, we have $\gamma_{i+1}=-2\gamma_i$ for $i=k-t,\ldots,k-1$.*

Proof. Denote $g = f \oplus \mathrm{PARITY}$ and let $G\colon \{0,\dots,k\} \to \{0,1\}$ be its univariate representation. Since we assume that g is t-null, it follows from Lemma 3.4 that $\mathrm{bias}(G|_{(\ell,0)}) = \mathrm{bias}(G|_{(\ell+1,0)})$ for $\ell = 0,\dots,t-1$. Therefore,

$$\begin{aligned}\frac{1}{2^{k-\ell}} \cdot \sum_{i=0}^{k-\ell} \binom{k-\ell}{i} \cdot (-1)^{G(i)} &= \frac{1}{2^{k-\ell}} \cdot \sum_{i=0}^{k-\ell} \binom{k-\ell}{i} \cdot (1 - 2G(i)) \\ &= 1 - 2\,\mathrm{bias}(G|_{(\ell,0)}) = 1 - 2\,\mathrm{bias}(G|_{(\ell+1,0)}) \\ &= \frac{1}{2^{k-\ell-1}} \cdot \sum_{i=0}^{k-\ell-1} \binom{k-\ell-1}{i} \cdot (1 - 2G(i)) \\ &= \frac{1}{2^{k-\ell-1}} \cdot \sum_{i=0}^{k-\ell-1} \binom{k-\ell-1}{i} \cdot (-1)^{G(i)}.\end{aligned}$$

Multiplying both sides by $2^{k-\ell}$ and using the fact that $(-1)^{G(i)} = (-1)^{F(i)} \cdot (-1)^i$ we get

$$\sum_{i=0}^{k-\ell} \binom{k-\ell}{i} \cdot (-1)^{F(i)} \cdot (-1)^i = 2 \cdot \sum_{i=0}^{k-\ell-1} \binom{k-\ell-1}{i} \cdot (-1)^{F(i)} \cdot (-1)^i.$$

Since $F(i) = (1-(-1)^{F(i)})/2$, and $\sum_{i=0}^{d} \binom{d}{i} \cdot (-1)^i = 0$, it follows that

$$\sum_{i=0}^{k-\ell} \binom{k-\ell}{i} \cdot F(i) \cdot (-1)^i = 2 \cdot \sum_{i=0}^{k-\ell-1} \binom{k-\ell-1}{i} \cdot F(i) \cdot (-1)^i. \tag{11}$$

By Lemma 2.5 we have $(-1)^d \gamma_d = \sum_{i=0}^{d} \binom{d}{i} \cdot F(i) \cdot (-1)^i$. Hence, Equation (11) is equivalent to

$$(-1)^{k-\ell} \cdot \gamma_{k-\ell} = 2 \cdot (-1)^{k-\ell-1} \cdot \gamma_{k-\ell-1},$$

i.e. $\gamma_{k-\ell} = -2 \cdot \gamma_{k-\ell-1}$. The claim now follows as this holds for every $\ell = 0,\dots,t-1$. ∎

We now show the connection between t-null functions and polynomials to $\{-1,0,1\}$.

Theorem 4.2. *If $f \oplus \mathrm{PARITY}$ is t-null then the interpolation polynomial of $F(|x|+2) - F(|x|)$ on the range $\{0,1,\dots,k-2\}$ is of degree smaller than $k-t-1$.*

Proof. Let $G(|x|) = F(|x|+2) - F(|x|)$. We compute G's representation in the Newton basis using F's representation. As before, denote $F(|x|) = \sum_{i=0}^{k} \gamma_i \binom{|x|}{i}$. The binomial formula $\binom{|x|+2}{i} = \binom{|x|}{i-2} + 2 \cdot \binom{|x|}{i-1} + \binom{|x|}{i}$ gives

$$\begin{aligned}
G(|x|) &= \sum_{i=0}^{k} \gamma_i \cdot \left[\binom{|x|+2}{i} - \binom{|x|}{i} \right] \\
&= \gamma_0 \cdot 0 + \gamma_1 \cdot 2 + \sum_{i=2}^{k} \gamma_i \cdot \left[\binom{|x|}{i-2} + 2 \cdot \binom{|x|}{i-1} + \binom{|x|}{i} - \binom{|x|}{i} \right] \\
&= \gamma_1 \cdot 2 + \sum_{i=2}^{k} \gamma_i \cdot \left[\binom{|x|}{i-2} + 2 \cdot \binom{|x|}{i-1} \right] \\
&= \gamma_1 \cdot 2 + \binom{|x|}{0} \cdot \gamma_2 + \sum_{i=1}^{k-2} \left[(\gamma_{i+2} + 2 \cdot \gamma_{i+1}) \cdot \binom{|x|}{i} \right] \\
&\quad + 2 \cdot \gamma_k \binom{|x|}{k-1} \\
&=^{(*)} \gamma_1 \cdot 2 + \gamma_2 \cdot \binom{|x|}{0} + \sum_{i=1}^{k-t-2} \left[(\gamma_{i+2} + 2 \cdot \gamma_{i+1}) \cdot \binom{|x|}{i} \right] \\
&\quad + 2 \cdot \gamma_k \binom{|x|}{k-1},
\end{aligned}$$

where equality $(*)$ follows from Lemma 4.1, as F is t-null. Let $H(|x|)$ be the interpolation polynomial at the points $\{0,1,\ldots,k-2\}$. In other words, $H(i) = G(i)$ for $i = 0,1,\ldots,k-2$, and $\deg(H) \le k-2$. By Lemma 2.5 the coefficients of $\binom{|x|}{i}$ in G and H (for $i = 0,1,\ldots,k-2$) are equal as they depend on the same set of values. Since $\deg(H) \le k-2$ it must be the case that

$$\begin{aligned}
H(|x|) &= G(|x|) - 2 \cdot \gamma_k \cdot \binom{|x|}{k-1} \\
&= \gamma_1 \cdot 2 + \gamma_2 \cdot \binom{|x|}{0} + \sum_{i=1}^{k-t-2} (\gamma_{i+2} + 2 \cdot \gamma_{i+1}) \cdot \binom{|x|}{i}.
\end{aligned}$$

In particular, $\deg(H) < k-t-1$. Finally, as $H(i)$ and $G(i)$ agree on $i = 0,1,\ldots,k-2$ and $G(i) = F(i+2) - F(i)$, we have that H maps $\{0,\ldots,k-2\}$ to $\{-1,0,1\}$. ∎

Theorem 1.4 is now an easy corollary. We repeat the statement of the theorem.

Theorem (Theorem 1.4, restated). *If the degree of any non-constant polynomial $H\colon \{0,\dots,k-2\}\to\{-1,0,1\}$ is at least $k-t$, then every non-linear symmetric function $f\colon \{0,1\}^k\to\{0,1\}$ must satisfy $\hat{f}(S)\neq 0$ for some non-empty S of size $|S|<t$.*

Proof. Assume by contradiction that there exists a non-linear symmetric function f which is $(t-1)$-null. Let $g=f\oplus\text{PARITY}$. Hence, $f=g\oplus\text{PARITY}$. Theorem 4.2 implies that the degree of the polynomial agreeing with $G(y+2)-G(y)$ on $\{0,\dots,k-2\}$ is smaller than $k-t$. By our assumption, it follows that $G(y+2)-G(y)$ is constant on $\{0,\dots,k-2\}$. Since G only attains the values 0 and 1, it must be the case that $G(y+2)-G(y)=0$ on $\{0,\dots,k-2\}$ (assuming[2] $k\geq 4$). Hence, G is equal to some constant on all the even elements in $\{0,\dots,k\}$ and to some (possibly different) constant on all the odd elements there. From the definition of g it follows that f has the same property. This can only happen if f is linear, which contradicts our assumption. ∎

Strengthening a conjecture of von zur Gathen and Roche [7], Mossel et al. [10] conjectured that any non-linear symmetric function must have a Fourier coefficient of size $O(1)$. Theorem 1.4 thus suggests a possible approach for proving this conjecture. We note, however, that obtaining better bounds even when the range of H is $\{0,1\}$ is still open. The best bounds that are currently known are $\deg(H)\geq k-\Gamma(k)$ when $H\colon \{0,\dots,k-2\}\to\{0,1\}$ [7], and $\deg(H)\geq k-O(\frac{k}{\log\log k})$ when $H\colon \{0,\dots,k-2\}\to\{-1,0,1\}$ [6].

5. Conclusion

In this paper we proved an upper bound on the minimal size of a (non-empty) set S such that $\hat{f}(S)\neq 0$ for any non-linear symmetric function f. We observed that the problem is related to the problem of giving a lower bound on the degree of any non-constant symmetric function. In Section 4 we saw the connection to lower bounding the degrees of symmetric functions taking values in $\{0,1,2\}$. To make the connection between the problems clearer we note that Theorem 1.4 implies that the problem of lower bounding the degree

[2] When $k=3$ the claim follows by inspection, noticing that $t=2$ and that any symmetric function on 3 bits has a nonzero Fourier coefficient of degree 1. For $k=2$, while the assumption is meaningless, it is easy to verify that a non-linear f has a degree 1 nonzero Fourier coefficient.

of functions into $\{0,1,2\}$ is at least as hard as proving an upper bound on the size of the first (non-empty) non-zero Fourier coefficient. The latter question in turn, is at least as difficult as proving a lower bound on the degree of any symmetric function (into $\{0,1\}$) as discussed in the introduction (since this problem is just a specialization to the case $\hat{f}(\emptyset)=0$).

Another interesting question is to get rid of the need to use number theory (i.e. Theorem 2.3). It is clear that new techniques are required as all current techniques rely on modular analysis which needs to assume the existence of primes in a certain range.

Acknowledgement. We thank the anonymous referee for bringing [12, 15] to our attention and for comments that improved the presentation of the results.

References

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

[2] R. C. Baker, G. Harman and J. Pintz: The difference between consecutive primes, II, *Proceedings of the London Mathematical Society* **83** (2001), 532–562.

[3] A. Blum and P. Langley: Selection of relevant features and examples in machine learning, *Artif. Intell.* **97** (1997), 245–271.

[4] A. L. Blum: Relevant examples and relevant features: Thoughts from computational learning theory, in: *In AAAI Fall Symposium on 'Relevance'*, 1994.

[5] H. Cramér: On the order of magnitude of the difference between consecutive prime numbers, *Acta Arithmetica* **2** (1936), 23–46.

[6] G. Cohen, A. Shpilka and A. Tal: On the degree of univariate polynomials over the integers, in: *Innovations in Theoretical Computer Science (ITCS)*, 409–427, 2012.

[7] J. von zur Gathen and J. R. Roche: Polynomials with two values, *Combinatorica* **17** (1997), 345–362.

[8] M. N. Kolountzakis, R. J. Lipton, E. Markakis, A. Mehta and N. K. Vishnoi: On the Fourier spectrum of symmetric Boolean functions, *Combinatorica* **29** (2009), 363–387.

[9] D. E. Knuth: *The Art of Computer Programming, Volume III: Sorting and Searching*, Addison-Wesley, 1973.

[10] E. Mossel, R. O'Donnell and R. A. Servedio: Learning functions of k relevant variables, *J. Comput. Syst. Sci.* **69** (2004), 421–434.

[11] N. Nisan and M. Szegedy: On the degree of Boolean functions as real polynomials. *Computational Complexity* **4** (1994), 301–313.

[12] T. Siegenthaler: Correlation-immunity of nonlinear combining functions for cryptographic applications, *IEEE TIT* **30** (1984), 776–780.

[13] A. Shpilka and A. Tal: On the minimal fourier degree of symmetric boolean functions, in: *Proceedings of the 26th Annual IEEE Conference on Computational Complexit (CCC)*, 200–209, 2011.

[14] G. Valiant: Finding correlations in subquadratic time, with applications to learning parities and juntas, in: *FOCS*, 11–20, 2012.

[15] G. Z. Xiao and J. L. Massey: A spectral characterization of correlation-immune combining functions, *IEEE TIT* **34** (1988), 569–571.

Amir Shpilka

Faculty of Computer Science
Technion-Israel Institute of Technology
Haifa, Israel and Microsoft Research
Cambridge
shpilka@cs.technion.ac.il

Avishay Tal

Department of Computer Science
and Applied Mathematics
Weizmann Institute of Science
Rehovot, Israel
avishay.tal@weizmann.ac.il