On the Fourier Analysis of Boolean Functions

A. Bernasconi[*] B. Codenotti[†] J. Simon[‡]

IMC B4-97-03

Abstract

We study the Fourier representation of Boolean functions. The goal is to look at the *frequency domain* of Boolean functions to get complexity properties. Preliminary results indicate that this might be fruitful. In addition to presenting new results, we review some of the most significant work on the subject.

[*]Istituto di Matematica Computazionale, Consiglio Nazionale delle Ricerche, and Dipartimento di Informatica, Pisa (Italy).

[†]Istituto di Matematica Computazionale, Consiglio Nazionale delle Ricerche, Pisa (Italy). e-mail: codenotti@iei.pi.cnr.it

[‡]Department of Computer Science, The University of Chicago. Portions of this work were done while visiting IEI-CNR in Pisa, sponsored by a grant from CNR.

1 Introduction

The Fourier transform of a Boolean function is an invertible linear mapping of the values of the function onto a set of coefficients, known as Fourier coefficients. This transformation is such that the Fourier coefficients contain information about the *regularities* of the function, and thus about its computational complexity.

While Fourier transforms of Boolean functions were used since the 60's, and several beautiful applications to computational complexity were obtained recently [LMN 93], we know of no attempt to study systematically the Fourier transform, with the aim of gaining new insights for the analysis of Boolean functions.

Indeed we would expect several potential benefits from such a study:

- definition of interesting new complexity classes defined by properties of the Fourier transforms of the functions in the class;

- new lower bounds proofs;

- a better understanding of properties of Boolean circuits.

In short, we expect benefits similar to the insights gained by electrical engineers when they think in terms of the "frequency domain".

This paper is organized as follows.
In Section 2 we give some background on abstract harmonic analysis and study the Fourier representation of Boolean functions. In Section 3 we give an interesting interpretation of the harmonic analysis on the hypercube in terms of the *Laplacian Matrix.* This leads to a natural view of the connections between the notions of *average* and *maximal sensitivity* on one side, and of Fourier coefficients on the other one. For the set of all Boolean functions of a given number of variables, we also give precise estimates for the distribution of the average sensitivity.
In Section 4 we relate the Fourier coefficients to the *influences* of sets of variables, and in Section 5 we show how to compute the Fourier coefficients and estimate the sensitivity of symmetric, monotone and read-once functions. In Section 6 we present some applications of Fourier Analysis to Circuit Complexity. We first review a result by Linial et al. (see [LMN 93]) about a spectral characterization of AC^0 functions, and then present some new results: using certain connections between *average sensitivity* and sparsity of languages, we characterize spectrum and sensitivity of symmetric AC^0 functions; we also define complexity classes characterized by functions that have most of their power spectrum in a relatively narrow contiguous band. Such classes generalize the well-known class AC^0 (it is well known [LMN 93] that AC^0 has such a property). We show that, like AC^0, they are learnable in quasi-polynomial time, under the uniform distribution. In fact, we prove that any class of Boolean functions with the property that most of its power spectrum is concentrated in a known set of coefficients, of at most polylogarithmic cardinality, is approximately learnable under the uniform distribution. We also derive a version of Khrapchenko's theorem expressed in terms of Fourier coefficients.

1.1 History

Fourier techniques in the Boolean setting were well known in the late 60's. They were used as a tool for analysis and synthesis of switching circuits, especially as a tool for sensitivity analysis and error detection. A review article of 1971 [Leh 71] illustrates the applications of Fourier analysis to the theory of switching circuits. Chapter 2 in [SHB 68] is dedicated to the description of some mathematical background on error detection in digital machines. In particular, the *Boolean difference* function - which is the difference between the value of a function at a given argument w and the value at an argument which is a Hamming neighbor of w - is used to analyze error propagation. Although without real justification, Karpovsky [Kar 76] has proposed to use the number of non-vanishing Fourier coefficients of a Boolean function f as a measure of its complexity.

While most of these results were forgotten, Fourier techniques reappeared in the 80's. Hurst et al. [HMM 82] relate the circuit complexity of a Boolean function to its power spectrum coefficients. Brandman et al. [BOH 90] establish a relationship between the Fourier coefficients of a Boolean function f and (i) the average size of any decision tree for f; (ii) the minimum number of $\wedge$ gates in a circuit computing f according to its disjunctive normal form. Kahan et al. [KKL 88] find connections between the influence of variables and harmonic analysis and use theorems on influence to prove results on rapidly mixing Markov chains. In addition, they relate the average sensitivity of functions to their Fourier coefficients. However this relation was already implicit in the work of [HMM 82]. Ben-Or and Linial [BL 89] study collective coin flipping, where the collective coin is viewed as a Boolean function. In this case, measuring the influence of a subset of variables onto the value of the function corresponds to measuring how the collective coin is sensitive to the presence of faults and the goal is to find Boolean functions on which the influence of each variable is as small as possible, to prevent a small subset of variables, e.g. the set of faulty processors, taking control of the collective coin. Linial et al. [LMN 93] take advantage of the relation between the average sensitivity of Boolean functions and their Fourier transform to prove several facts, e.g., that sets in AC^0 have low average sensitivity. Bruck [Bru 90] and Bruck and Smolensky [BS 92] use abstract harmonic analysis to derive necessary and sufficient conditions for a Boolean function to be a polynomial threshold. Finally, in recent years, several results in Learning Theory have been obtained by applying Fourier-based techniques (see [Man 94] and [Jac 95] for a comprehensive and exhaustive survey on this topic).

1.2 Definitions

Unless otherwise specified, the indexing of vectors and matrices starts from 0 rather than 1. The symbol e_1 denotes the first column of the identity matrix. A^T denotes the transpose of a matrix A. $\rho(B)$ denotes the spectral radius of a matrix B, i.e. the largest of the absolute values of the eigenvalues of B. The notation $\|x\|$ ($\|B\|$) without any subscript stands for the L_2-norm of a vector x (matrix B). The subscript 1 is used to specify the L_1-norm of vectors and matrices. All the logarithms are to the base 2. The notation *polylogn* stands for a function growing like a polynomial in the logarithm of n. Given a Boolean function f on n binary variables, we will use different kinds of notation: the *classical notation*, where the input string is given by n binary variables; the *set notation*, based on the correspondence between the set $\{0,1\}^n$ and and the power set of $\{1,2,\ldots,n\}$; the 2^n tuple *vector representation* $f = (f_0\, f_1 \ldots f_{2^n-1})$, where $f_i = f(x(i))$ and $x(i)$ is the binary expansion of i. If x and y are two binary strings of the same length, then $d(x,y)$ and $x \oplus y$ denote their Hamming distance and the string obtained by computing the *exclusive or* of the bits of x and y, respectively. $|x|$ denotes the number of ones in a binary string x.

2 Fourier Transform of Boolean Functions

We give some background on abstract harmonic analysis. Our main sources are [Leh 71] and [Loo 53].

We consider Boolean functions as 0-1 valued real functions defined on the domain $\{0,1\}^n$. They are a vector space of dimension 2^n, and the set of functions $\{f_x(y) = (1 \text{ iff } x = y)\}$, where x ranges over $\{0,1\}^n$ is a basis. Another basis is given by the functions $\{g_S(x) = \sum_{i \in S} x_i\}$, where the sum is modulo 2. The Fourier coefficients of f are the coefficients of f in this basis.

More precisely, consider the space $\mathcal{F}$ of all the two-valued functions on $\{0,1\}^n$. The domain of $\mathcal{F}$ is a locally compact Abelian group and the elements of its range, i.e. 0 and 1, can be added and multiplied as complex numbers. The above properties allow one to analyze $\mathcal{F}$ by using tools from harmonic analysis. This means that it is possible to construct an orthogonal basis set of Fourier transform kernel functions for $\mathcal{F}$. The kernel functions of the Fourier transform are defined in terms of a group homomorphism from $\{0,1\}^n$ to the direct product of n copies of the multiplicative subgroup $\{\pm 1\}$ on the unit circle of the complex plane. The functions $Q_w(x) = (-1)^{w_1 x_1}(-1)^{w_2 x_2} \ldots (-1)^{w_n x_n} = (-1)^{w^T x}$ are known as *group characters* or Fourier transform kernel functions [Lit 40]. The set of functions $\{Q_w | w \in \{0,1\}^n\}$ is an orthogonal basis for $\mathcal{F}$.

We can now define the *Abstract Fourier Transform* of a Boolean function f as the rational valued function f^* which defines the coordinates of f with respect to the basis $\{Q_w(x), w \in \{0,1\}^n\}$, i.e., $f^*(w) = 2^{-n} \sum_x Q_w(x) f(x)$. Then $f(x) = \sum_w Q_w(x) f^*(w)$ is the Fourier expansion of f.

Using the binary 2^n-tuple representation for the functions f and f^*, and considering the natural ordering of the n-tuples x and w, one can derive a convenient matrix formulation for the transform pair. Let us consider a $2^n \times 2^n$ matrix H_n whose (i,j)-th entry h_{ij} satisfies $h_{ij} = (-1)^{x(i)^T x(j)}$, where $x(i)^T x(j)$ denotes the inner product of the binary expansions of i and j. If $f = [f_0\, f_1 \ldots f_{2^n-1}]^T$ and $f^* = [f_0^*\, f_1^* \ldots f_{2^n-1}^*]^T$, then, from the fact that $H_n^{-1} = 2^{-n} H_n$, we get $f = H_n f^*$ and $f^* = 2^{-n} H_n f$.

Note that the matrix H_n is the Hadamard symmetric transform matrix [Leh 71] and can be recursively defined as

$$H_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H_n = \begin{pmatrix} H_{n-1} & H_{n-1} \\ H_{n-1} & -H_{n-1} \end{pmatrix}.$$

2.1 Interpretation of the Fourier coefficients

We now give an interpretation of the Fourier coefficients, making use of the set notation. We first need some definitions.

Definition 1 *The* order *of a Fourier coefficients $f^*(S)$ is the cardinality of the subset S.*

Definition 2 *The* degree *of a Boolean function f is the maximum order of the non-zero Fourier coefficients, i.e.*

$$deg(f) = \max\{\ |S| \mid f^*(S) \neq 0\ \}\,.$$

Note that this definition of degree coincides with the degree related to the polynomial representation of Boolean functions over the reals (see [B93]).

In the following, we show that the Fourier coefficients measure the correlation under the uniform distribution between the function and parities of certain subsets of variables.

- The coefficient $f^*(\emptyset)$ is simply the probability that f takes the value 1. Indeed we have $f^*(\emptyset) = \frac{1}{2^n}\sum_w f(w)$.
- The first order coefficients $f^*(i)$ measure the correlation of the function f with its i-th variable. It is easy to prove that

$$f^*(\{i\}) = \frac{1}{2} - \Pr[f(w) = w_i]\,.$$

  There is no correlation if $f^*(\{i\}) = 0$ and maximum correlation if $|f^*(\{i\})| = \frac{1}{2}$. The sign of the coefficient indicates if the correlation is actually with the variable w_i ($f^*(\{i\}) = -\frac{1}{2}$) or with its complement $\neg w_i$ ($f^*(\{i\}) = \frac{1}{2}$).
- The Fourier coefficients of order k, $f^*(S)$, $|S| = k$, measure the correlation between the function and the parity of those variables w_j such that $j \in S$:

$$f^*(S) = \frac{1}{2} - \Pr[f(w) = \bigoplus_{j\in S} w_j]\,.$$

  Again, there is no correlation if $f^*(S) = 0$, and maximum correlation if $|f^*(S)| = \frac{1}{2}$. The sign of the coefficient indicates if the correlation is actually with the parity ($f^*(S) = -\frac{1}{2}$) or with its complement ($f^*(S) = \frac{1}{2}$).

3 Harmonic analysis on the hypercube and sensitivity

We now give an interpretation of the harmonic analysis on the hypercube in terms of its *Laplacian Matrix*, and will use it below, to describe the interplay between the Fourier coefficients and the *sensitivity* of Boolean functions.

First of all, recall that the *n-dimensional hypercube* consists of 2^n vertices $\{0,1\}^n$, and that there is an edge between two vertices if and only if they differ in exactly one coordinate. Since each vertex has n adjacent edges, the n-dimensional hypercube is an n-regular graph. We now define its *adjacency* and *Laplacian* matrices.

Definition 3

- *The* $2^n \times 2^n$ adjacency matrix *of the* n*-dimensional hypercube is recursively defined as:*
$$A_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad A_n = \begin{pmatrix} A_{n-1} & I_{n-1} \\ I_{n-1} & A_{n-1} \end{pmatrix},$$
where I_k *denotes the* $2^k \times 2^k$ *identity matrix.*

- *The* $2^n \times 2^n$ Laplacian matrix *of the* n*-dimensional hypercube is defined as:*
$$\mathcal{L}_n = nI_n - A_n .$$

The following theorems show the intimate relationship between $\mathcal{L}_n$ and the Hadamard matrix H_n. For any integer $0 \leq i \leq 2^n - 1$, we denote by $|i|$ the number of ones in its binary expansion $x(i) \in \{0,1\}^n$.

Theorem 1 *For any* n*,* $\mathcal{L}_n = \frac{1}{2^{n-1}} H_n D_n H_n$ *, where* H_n *is the Hadamard matrix and* D_n *is the* $2^n \times 2^n$ *diagonal matrix whose* (i,i)*-entry is equal to* $|i|$*, for* $0 \leq i \leq 2^n - 1$*.*

Proof. By induction on n.

Base

By direct inspection.

Induction step

We consider the $2^{n+1} \times 2^{n+1}$ Laplacian matrix $\mathcal{L}_{n+1}$. Since
$$\mathcal{L}_{n+1} = \begin{pmatrix} \mathcal{L}_n + I_n & -I_n \\ -I_n & \mathcal{L}_n + I \end{pmatrix} \quad \text{and} \quad D_{n+1} = \begin{pmatrix} D_n & 0 \\ 0 & D_n + I_n \end{pmatrix},$$
the theorem immediately follows by induction hypothesis, i.e.,
$$\begin{aligned} \frac{1}{2^n} H_{n+1} D_{n+1} H_{n+1} &= \frac{1}{2^n} \begin{pmatrix} 2H_n D_n H_n + H_n H_n & -H_n H_n \\ -H_n H_n & 2H_n D_n H_n + H_n H_n \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{L}_n + I_n & -I_n \\ -I_n & \mathcal{L}_n + I_n \end{pmatrix}. \end{aligned}$$

■

Corollary 2 H_n *is the matrix of the eigenvectors of the Laplacian matrix (as well as of the adjacency matrix) of the* n*-dimensional hypercube.*

Proof. Easily follows from theorem 1, since $H_n^{-1} = \frac{1}{2^n} H_n$.

■

The eigenvalues of the Laplacian matrix are given by $\lambda(i) = 2|i|$, for any $0 \leq i \leq 2^n - 1$.

The *sensitivity* of a Boolean function f measures if it is likely that $f(w) \neq f(\hat{w})$, if w and $\hat{w}$ are Hamming neighbors. More precisely, we have the following definitions.

Definition 4

- *The* sensitivity $s_f(w)$ *of* f *on a string* $w \in \{0,1\}^n$ *is the number of Hamming neighbors* $\hat{w}$ *of* w *such that* $f(w) \neq f(\hat{w})$.

- *The* average sensitivity *of* f, $s(f)$, *is the average of* $s_f(w)$ *over all* $w \in \{0,1\}^n$. $s(f)$ *can also be defined as the sum of the influences of all the variables on* f *(see def. 5).*

- *The* maximal sensitivity *of* f *is defined as* $s_{max}(f) = \max_w\{s_f(w)\}$.

(Sometimes we will use the terminology *sensitivity of a set* as a shortcut for *sensitivity of the characteristic function of the elements of length* n *of a set.*)

Lemma 3 shows how it is possible to express the sensitivity of Boolean functions in terms of the Laplacian matrix. Let f be a Boolean function depending on n variables and $\tilde{s}$ the vector defined as $\tilde{s} = \mathcal{L}_n f$.

Lemma 3 *For any* $0 \leq i \leq 2^n - 1$, *the absolute value of the* i^{th} *component of* $\tilde{s}$ *is equal to the sensitivity of* f *on the string* $x(i)$*; more precisely:*

$$\tilde{s}_i = -(-1)^{f(x(i))}\ s_f(x(i))\,.$$

Proof. Let us consider the i^{th} entry of the vector $\tilde{s} = \mathcal{L}_n f$. We have

$$\tilde{s}_i = nf(x(i)) - (A_n f)_i = nf(x(i)) - \sum_{|\mu|=1} f(x(i) \oplus \mu)\,.$$

Thus we obtain

$$\tilde{s}_i = \begin{cases} n - \sum_{|\mu|=1} f(x(i) \oplus \mu) = s_f(x(i)) & \text{if } f(x(i)) = 1\,, \\ -\sum_{|\mu|=1} f(x(i) \oplus \mu) = -s_f(x(i)) & \text{if } f(x(i)) = 0\,. \end{cases}$$

■

Corollary 4 *For a Boolean function* f, *we have:*

(*i*) $s_{max}(f) = \|\mathcal{L}_n f\|_\infty$.

(*ii*) $s(f) = \frac{1}{2^n}\|\mathcal{L}_n f\|_1$.

Proof. Immediate from lemma 3.

■

We now use these results to analyze the connections between sensitivity and Fourier coefficients: we find a relation between the maximal sensitivity and the Fourier coefficients of a function and obtain a new proof for the already known relation bertween the average sensitivity and the Fourier spectrum (see [HMM 82] for the result on the average sensitivity $s(f)$).

Lemma 5

1. $s_{max}(f) = \max_w \left|2 \sum_v (-1)^{v^T w} |v| f^*(v)\right|$.
2. $s(f) = \frac{1}{2^{n-1}} f^T \mathcal{L}_n f = 4 \sum_w |w| (f^*(w))^2$.

Proof.

1. From theorem 1 and from the definition of Fourier transform of a Boolean function, it follows that
$$\tilde{s} = \mathcal{L}_n f = \frac{2}{2^n} H_n D_n H_n f = 2 H_n D_n f^* ,$$
which yields
$$|\tilde{s}_i| = 2 \left| \sum_v (-1)^{x(i)^T v} |v| f^*(v) \right| .$$
The proof thus follows from the equality $|\tilde{s}_i| = s_f(x(i))$.

2. Let u denote the vector whose entries are all equal to 1. Lemma 3 yields
$$s(f) = \frac{1}{2^n} (2f - u)^T \mathcal{L}_n f = \frac{1}{2^{n-1}} f^T \mathcal{L}_n f ,$$
where the last equality follows since $u^T \mathcal{L}_n f = 0$, as one can easily check. Then, from theorem 1 we can derive
$$s(f) = \frac{1}{2^{n-1}} f^T \mathcal{L}_n f = \frac{1}{2^{2(n-1)}} f^T H_n D_n H_n f = 4 (f^*)^T D_n (f^*) = 4 \sum_w |w| (f^*(w))^2 .$$

■

3.1 Distribution of the average sensitivity

We analyze the distribution of the average sensitivity of all Boolean functions defined on $\{0,1\}^n$ and we give an exact evaluation of its expected value and variance. Then we use Chebyshev's inequality to find an upper bound on the number of Boolean functions in AC^0. First of all, note that, for any function f, $0 \le s(f) \le n$ and that $s(f)$ can assume only rational values.

Theorem 6 *The average sensitivity of all 2^{2^n} Boolean functions of n variables has expected value $\frac{n}{2}$ and variance $\frac{n}{2^{n+2}}$.*

Proof. The first claim easily follows from the linearity of the expected value. In fact:
$$\begin{aligned}
E[s(f)] &= E\left[\frac{1}{2^n} \sum_w s_f(w)\right] = \frac{1}{2^n} \sum_w E[s_f(w)] \\
&= \frac{1}{2^n} \sum_w \sum_{|u|=1} E[f(w) + f(w \oplus u) - 2 f(w) f(w \oplus u)] \\
&= \frac{1}{2^n} \sum_w \sum_{|u|=1} \left(E[f(w)] + E[f(w \oplus u)] - 2E[f(w) f(w \oplus u)]\right) \\
&= \frac{1}{2^n} \sum_w \sum_{|u|=1} \left(\frac{1}{2} + \frac{1}{2} - 2 \cdot \frac{1}{4}\right) = \frac{n}{2} .
\end{aligned}$$

For the second claim, from the definition of variance we get

$$V[s(f)] = E[s^2(f)] - E^2[s(f)].$$

Then, we need evaluate the expected value of $s^2(f)$. First of all, we have

$$\begin{aligned} E[s^2(f)] &= E\left[\frac{1}{2^{2n}}\sum_w \sum_{w'} s_f(w)s_f(w')\right] \\ &= \frac{1}{2^{2n}} E\left[\sum_w s_w^2(f) + \sum_{w\neq w'} s_f(w)s_f(w')\right] \\ &= \frac{1}{2^{2n}}\left(\sum_w E[s_w^2(f)] + \sum_{w\neq w'} E[s_f(w)s_f(w')]\right). \end{aligned}$$

Since

$$E[S_w^2(F)] = \sum_{I=1}^{n} E[d_i^2(w)] + \sum_{i\neq j} E[d_i(w)d_j(w)],$$

where $d_i(w) = f(w) + f(w \oplus i) - 2f(w)f(w\oplus i)$ and i is a string with the i-th bit equal to 1 and the others equal to 0, by multiplying $d_i(w)$ and $d_j(w)$ and evaluating the expected values, we obtain

$$E[s_w^2(f)] = \frac{n^2}{4} + \frac{n}{4}.$$

In the same way, we find

$$E[s_f(w)s_f(w')] = \sum_{i,j} E[d_i(w)d_j(w')] = \frac{n^2}{4}.$$

Note that $E[d_i(w)d_j(w')]$ turns out to be equal to $1/4$ even if w and w' share a Hamming neighbor, i.e. $w \oplus i = w' \oplus j$ for some i and j, and even if w is a Hamming neighbor of w' , i.e. $w = w' \oplus j$ for some j (and *vice-versa*). Finally, we get

$$E[s^2(f)] = \frac{1}{2^{2n}}\left[2^n\left(\frac{n^2}{4}+\frac{n}{4}\right) + \left(2^{2n}-2^n\right)\frac{n^2}{4}\right] = \frac{n}{2^{n+2}} + \frac{n^2}{4}$$

and the thesis follows from $E[s(f)] = \frac{n}{2}$.

■

Lemma 7 *The distribution of the average sensitivity of all the 2^{2^n} Boolean functions of n variables is symmetric with respect to the expectation, i.e.*

$$\#\{\ f:\{0,1\}^n \to \{0,1\} \mid s(f) = k\ \}| = \#\{\ f:\{0,1\}^n \to \{0,1\} \mid s(f) = n-k\ \}|,$$

Proof. We show that for any Boolean function f, with average sensitivity $s(f)$, there exists a function g s.t. $s(g) = n - s(f)$. We define g in the following way. For all $w \in \{0,1\}^n$,

$$g(w) = \begin{cases} 1 - f(w) & \text{if } |w| \text{ is odd}, \\ f(w) & \text{if } |w| \text{ is even}. \end{cases}$$

The function g is such that, for all w, $s_g(w) = n - s_f(w)$. In fact for all the strings u, $|u| = 1$, w and $w \oplus u$ have opposite parities, thus we obtain

$$\begin{aligned} s_g(w) &= \sum_{|u|=1} |g(w) - g(w \oplus u)| \\ &= \sum_{|u|=1} (1 - |f(w) - f(w \oplus u)|) = n - s_f(w). \end{aligned}$$

Hence

$$s(g) = \frac{1}{2^n} \sum_w s_g(w) = n - s(f).$$

■

Note that from the two constant functions $f(w) = 0$ and $f'(w) = 1 - f(w) = 1$ for all w, which are the only two functions with average sensitivity equal to 0, we can use the proof of theorem 6 to get the two functions with the maximum average sensitivity, i.e., the parity and its complement. Theorem 8 is an application of Chebyshev's inequality, $\Pr\ \{|s(f) - E[s(f)]| \geq \varepsilon\} \leq \frac{V[s(f)]}{\varepsilon^2}$, to the distribution of average sensitivity.

Theorem 8 *The number of Boolean functions on n variables with $s(f) < k$, for $k < \frac{n}{2}$, is* $O\left(\frac{2^{2^n}}{n2^n\left(1-\frac{2k}{n}\right)^2}\right)$.

Proof. Follows from the Chebyshev's inequality, with $\varepsilon = \frac{n}{2} - k$.

■

Since Boolean functions in AC^0 have average sensitivity $s(f) \leq \log^{O(1)} n$ (cfr. [LMN 93]), we can use theorem 8 to prove that $\frac{2^{2^n}}{n2^n}\left(1 + \frac{4\log^{O(1)} n}{n}\right)$ is an upper bound on the number of functions in AC^0.

4 Fourier coefficients and Influences of variables

We now describe the links between harmonic analysis and the notions of *influences* of sets of variables. Let $\mathcal{A}$ be a set of variables.

Definition 5 *The* Influence *Of $\mathcal{A}$ On F, Denoted By $I_f(\mathcal{A})$, is the probability that f remains undetermined as long as the variables in $\mathcal{A}$ are not assigned values and the other variables are assigned at random according to the uniform distribution.*

The relations between influences and Fourier coefficients are clear when the function is monotone and only for first order coefficients. It was observed in [KKL 88] that for any monotone function f, and any variable x_i

$$I_f(x_i) = 2\,|f^*(\{i\})|\,.$$

This also holds for *locally monotone* Boolean function (see [GL 94]), i.e. functions that are monotonic increasing or decreasing in each variable. On the other hand, for general Boolean functions we get

$$I_f(x_i) \geq 2\,|f^*(\{i\})|\,.$$

In [GL 94] it was also proven that for any function, and any variable x_i

$$I_f(x_i) \geq 2 \max_{\{I : i \in I\}} \{|f^*(I)|\}\,.$$

Note that, from this result, it follows that for locally monotone functions, the linear Fourier coefficients are the largest in absolute value.

We find exact relationships between influences of subsets of variables and Fourier coefficients of any order. For the sake of exposition we present first the case of first order coefficients and then the general case.

4.1 First order coefficients

Let $w = x_1 x_2 \ldots x_n \in \{0,1\}^n$ and let $\tilde{w} = x_1 x_2 \ldots x_{i-1} \star x_{i+1} \ldots x_n \in \{0,1\}^n$ be the string obtained from w by replacing the i-th bit x_i with an undefined value $\star$. Let $S_i \subseteq \{0,1\}^n$ be the set $S_i \equiv \{w \in \{0,1\}^n \;\mid\; x_i = \star \;\Rightarrow\; f(\tilde{w}) = \star\}$. Let $S_i^{\oplus} \equiv \{w \in S_i \;\mid\; f(w) = x_i\}$ and $S_i^{\overline{\oplus}} \equiv \{w \in S_i \;\mid\; f(w) = \overline{x}_i\}$. Note that $S_i = S_i^{\oplus} \cup S_i^{\overline{\oplus}}$ and $S_i^{\oplus} \cap S_i^{\overline{\oplus}} = \emptyset$.

Definition 6 *The* influence of the variable x_i on the function f *is given by*

$$I_f(x_i) = \frac{1}{2^n}\;|S_i| = \frac{1}{2^n}\sum_w [1 - f(w) f(w \oplus i) - \overline{f(w)}\;\overline{f(w \oplus i)}].$$

Definition 7 *The* positive influence *and the* negative influence *of the variable x_i on f are given by $I_f^{\oplus}(x_i) = \frac{1}{2^n}|S_i^{\oplus}| = \frac{1}{2^n}\sum_{w \in S_i}(x_i f(w) + \overline{x}_i\;\overline{f(w)})$ and $I_f^{\overline{\oplus}}(x_i) = \frac{1}{2^n}|S_i^{\overline{\oplus}}| = \frac{1}{2^n}\sum_{w \in S_i}(x_i \overline{f(w)} + \overline{x}_i f(w))$.*

Observe that, for any $1 \leq i \leq n$, we have $I_f(x_i) = I_f^{\oplus}(x_i) + I_f^{\overline{\oplus}}(x_i)$.

Proposition 9 *For any $1 \leq i \leq n$ $\;f^*(\{i\}) = \frac{1}{2}\;[\; I_f^{\overline{\oplus}}(x_i) - I_f^{\oplus}(x_i)\;]$.*

The known relation between influences and the first order coefficients of locally monotone functions easily follows from proposition 9. In fact, for any locally monotone function and any variable x_i, one of the two components of the influence must vanish, and thus we obtain

$$|f^*(\{i\})| = \frac{1}{2}\; I_f(x_i)\,.$$

For the special case of monotone function, we get $I_f^{\overline{\oplus}}(x_i) = 0$, and then

$$f^*(\{i\}) = -\frac{1}{2}\; I_f^{\oplus}(x_i) = -\frac{1}{2}\; I_f(x_i)\,.$$

4.2 General case

Let $S_{i_1,\ldots,i_k} \subseteq \{0,1\}^n$ be a set defined as follows

$$S_{i_1,\ldots,i_k} \equiv \{w = x_1 x_2 \ldots x_n \in \{0,1\}^n \mid x_{i_1} = \star, \ldots, x_{i_k} = \star \;\Rightarrow\; f(\tilde{w}) = \star\},$$

where $\tilde{w}$ is the string obtained from w by replacing the bits $x_{i_1}, \ldots, x_{i_k}$ with undefined values $\star$. Then, let $S^{\oplus}_{i_1,\ldots,i_k}$ and $S^{\overline{\oplus}}_{i_1,\ldots,i_k}$ be the following sets

$$S^{\oplus}_{i_1,\ldots,i_k} \equiv \left\{ w = x_1 x_2 \ldots x_n \in S_{i_1,\ldots,i_k} \quad | \quad f(w) = \bigoplus_{l=i_1}^{i_k} x_l \right\},$$

$$S^{\overline{\oplus}}_{i_1,\ldots,i_k} \equiv \left\{ w = x_1 x_2 \ldots x_n \in S_{i_1,\ldots,i_k} \quad | \quad f(w) = \overline{\bigoplus_{l=i_1}^{i_k} x_l} \right\}.$$

Note that $S_{i_1,\ldots,i_k} = S^{\oplus}_{i_1,\ldots,i_k} \cup S^{\overline{\oplus}}_{i_1,\ldots,i_k}$ and $S^{\oplus}_{i} \cap S^{\overline{\oplus}}_{i_1,\ldots,i_k} = \emptyset$.

Definition 8 *The* influence of the variables $x_{i_1}, \ldots, x_{i_k}$ on the function f *is given by*

$$I_f(x_{i_1}, \ldots, x_{i_k}) = \frac{1}{2^n} \; |S_{i_1,\ldots,i_k}|.$$

Definition 9 *The* positive influence, $I^{\oplus}_f$, *and the* negative influence, $I^{\overline{\oplus}}_f$, *of the variables* $x_{i_1}, \ldots, x_{i_k}$ *on* f *are given by*

$$I^{\oplus}_f(x_{i_1}, \ldots, x_{i_k}) = \frac{1}{2^n} |S^{\oplus}_{i_1,\ldots,i_k}| = \frac{1}{2^n} \sum_{w \in S_{i_1,\ldots,i_k}} \left(\bigoplus_{l=i_1}^{i_k} x_l \; f(w) + \overline{\bigoplus_{l=i_1}^{i_k} x_l} \; \overline{f(w)} \right)$$

and

$$I^{\overline{\oplus}}_f(x_{i_1}, \ldots, x_{i_k}) = \frac{1}{2^n} |S^{\overline{\oplus}}_{i_1,\ldots,i_k}| = \frac{1}{2^n} \sum_{w \in S_{i_1,\ldots,i_k}} \left(\bigoplus_{l=i_1}^{i_k} x_l \; \overline{f(w)} + \overline{\bigoplus_{l=i_1}^{i_k} x_l} \; f(w) \right).$$

For all subsets of k variables $\{x_{i_1}, \ldots, x_{i_k}\}$ we get

$$I_f(x_{i_1}, \ldots, x_{i_k}) = I^{\oplus}_f(x_{i_1}, \ldots, x_{i_k}) + I^{\overline{\oplus}}_f(x_{i_1}, \ldots, x_{i_k}).$$

Proposition 10 *For all subsets of k variables $\{x_{i_1}, \ldots, x_{i_k}\}$*

$$f^*(\{i_1, \ldots, i_k\}) = \frac{1}{2} \; [\; I^{\overline{\oplus}}_f(x_{i_1}, \ldots, x_{i_k}) - I^{\oplus}_f(x_{i_1}, \ldots, x_{i_k}) \;].$$

Note that these results yield the relation $I_f(A) \geq 2\,|f^*(A)|$, for any Boolean function and any subset $A \subseteq \{1, 2, \ldots, n\}$.

5 Fourier Analysis of Symmetric, Monotone and Read Once functions

A basic computational problem involving Fourier coefficients is the following. Given a Boolean function f - say by the circuit C that computes it, obtain the Fourier coefficients of f - say by exhibiting an algorithm with input a set S (given by its characteristic vector) that outputs the corresponding Fourier coefficient of f, as a quotient of two integers. It is easy to see that the value of these integers is at most exponential, so they can be written in time polynomial in n. It is also clear from the previous paragraphs that the computation takes at most exponential time. It is an interesting open problem to find classes of Boolean functions for which we can compute the Fourier coefficients in time polynomial in the size of the circuits.

We know of three such classes: random functions (i.e. almost all Boolean functions), *symmetric* and *monotone* Boolean functions. It is unlikely that a general algorithm exists, since it would imply $P = PP$.

The existence of an algorithm for random functions is trivial: the coefficients can be computed in exponential time, which is polynomial in the size of the circuit (which is $\frac{2^n}{n}$).

As we have seen, the coefficient f_0^* is proportional to the number of assignments that give f the value 1, so computing this coefficient in P would yield a polynomial time algorithm for a PP-complete problem. On the other hand, it is not hard to see that every coefficient can be computed in polynomial time with a PP oracle.

We sketch below how to compute the Fourier coefficients and estimate the sensitivity of symmetric functions and monotone functions. Finally, we consider the special case of *Read-once formulae.*

Symmetric functions

The Fourier coefficients of symmetric functions can be expressed in terms of the Krawtchouk polynomials $K_k(x;n)$, $x \in R$, defined by

$$K_k(x;n) = \sum_{j=0}^{k} (-1)^j \binom{x}{j} \binom{n-x}{k-j} .$$

First of all, note that given a binary string w of length n, we get

$$f^*(w) = 2^{-n} \sum_{k=0}^{n} f_{2^k-1} \sum_{|x|=k} (-1)^{w^T x} .$$

Then, we observe that

$$\sum_{|x|=k} (-1)^{w^T x} = \sum_{i=0}^{k} (-1)^i \binom{n-|w|}{k-i} \binom{|w|}{i} ,$$

and finally we obtain

$$f^*(w) = 2^{-n} \sum_{k=0}^{n} f_{2^k-1} K_k(|w|;n) .$$

For the average sensitivity we get

$$s(f) = 4n \sum_{i=1}^{n} \binom{n-1}{i-1} \left(f^*_{2^i-1}\right)^2 .$$

Monotone functions

The n coefficients which are sufficient to determine the sensitivity [KKL 88] can be computed as $f^* = \hat{H}_n f$, where $\hat{H}_n$ is an $n \times 2^n$ matrix defined as

$$\hat{H}_1 = \begin{pmatrix} 1 & -1 \end{pmatrix}, \qquad \hat{H}_k = \begin{pmatrix} \hat{H}_{k-1} & \hat{H}_{k-1} \\ v_{k-1} & -v_{k-1} \end{pmatrix},$$

where $v_i = (1, 1 \ldots, 1)^T$. Then $s(f) = 2^{-n+1} \|\hat{H}_n f\|_1$.

Threshold functions

In the case of monotone and symmetric functions, i.e. *thresholds*, we get

$$f^*(w) = 2^{-n} \sum_{k=h}^{n} K_k(|w|; n) = -2^{-n} K_{h-1}(|w| - 1; n-1),$$

where h is the positive integer which defines the threshold, and where we used the following property of Krawtchouk polynomials

$$\sum_{k=0}^{l} K_k(x; n) = K_l(x - 1; n - 1).$$

In addition, note that for every $w',\ w \in \{0,1\}^n$ such that $|w'| = n - |w| + 1$ we have $|f^*(w)| = |f^*(w')|$.

Finally, the average sensitivity of h-threshold functions is given by

$$s(f) = \frac{h}{2^{n-1}} \binom{n}{h}.$$

An important threshold function is majority. For this function we have the following results. For n odd,

$$f^*(w) = \begin{cases} \frac{1}{2} & \text{if} \quad |w| \quad \text{is} \quad 0 \\ (-1)^{\frac{|w|+1}{2}} \frac{1}{2^n} \frac{\binom{|w|-1}{\frac{|w|-1}{2}} \binom{n-|w|}{\frac{n-|w|}{2}}}{\binom{\frac{n-1}{2}}{\frac{|w|-1}{2}}} & \text{if} \quad |w| \quad \text{is} \quad \text{odd} \\ 0 & \text{if} \quad |w| \quad \text{is} \quad \text{even.} \end{cases}$$

Moreover, using Stirling's approximation, we obtain, for $|w|$ odd,

$$|f^*(w)| \approx \frac{1}{\sqrt{\pi}} \left[\frac{(|w|-1)^{|w|-1} (n - |w|)^{n-|w|}}{(n-1)^n} \right]^{\frac{1}{2}}.$$

From this characterization it follows that $L_\infty(f)$ is inverse polynomially large, while $L_1(f)$ can not be bounded by any polynomial in n, where L_∞ and L_1 denote the norms associated with the spectrum of a function:

$$\begin{aligned} L_\infty(f) &= \max_w \{|f^*(w)|\} \\ L_1(f) &= \sum_w |f^*(w)|. \end{aligned}$$

Lemma 11 *Let δ be a positive constant such that $\frac{\sqrt{n}}{k\sqrt{k(n-k)}} \leq \frac{1}{k^{1+\delta}}$. Then*

$$\sum_{t<|S|\leq n-t} (f^*\{S\})^2 = \sum_{k=t}^{n-t} E_k(f) \leq \frac{1}{\pi\sqrt{2\pi}} \left[\frac{1}{\delta\ t^\delta} + o\left(\frac{1}{t^\delta}\right)\right].$$

Proof. The result follows from the fact that

$$\pi\sqrt{2\pi} \sum_{t<|S|\leq n-t} (f^*\{S\})^2 \leq \sum_{k=t}^{n-t} \frac{1}{k^{1+\delta}} \leq \frac{1}{t^{1+\delta}} + \int_t^{n-t} \frac{1}{x^{1+\delta}} dx \leq \frac{1}{t^{1+\delta}} + \frac{1}{\delta}\frac{1}{t^\delta}.$$

■

Functions computable by read-once formulae

We analyze the average sensitivity of Boolean functions computable by read-once formulas (read-once functions from now). Recall that a *formula* is a Boolean circuit of fan-out 1 and a *read-once formula* is a formula in which each variable appears only once. These functions are important, especially for "low" complexity classes, because every NC^1 function on n variables can be viewed as the projection of a read-once function with $n^{O(1)}$ variables.

To study read-once functions, we first introduce a more general notion of sensitivity and influence, which we call *on-line sensitivity* and *on-line influence*.

Definition 10 *The* on-line sensitivity $s_p(f)$ *of a Boolean function f is given by $s_p(f) = \sum_w p_w s_f(w)$, where p_w is the probability of occurrence of the argument w, and $s_f(w)$ is the sensitivity of f on w.*

Definition 11 *Let $\mathcal{A}$ be a set of variables. The* on-line influence *of $\mathcal{A}$ on f, $I_{f,p}(\mathcal{A})$, is the probability that f remains undetermined as long as the variables in $\mathcal{A}$ are not assigned values and the other variables are assigned according to a given probability distribution p.*

As for $s(f)$, one can evaluate $s_p(f)$ as the sum of the on-line influences of all the variables.

We are now able to characterize the sensitivity of read-once functions.

Lemma 12 *Let h be a Boolean function computable by a read-once formula and defined as $h(x) = f(g_1(w_1), g_2(w_2), \ldots, g_m(w_m))$, where $w_i = x_{1i}\, x_{2i} \ldots x_{k_i i}$.*

(a) *The influence of x_{ij} on h is $I_h(x_{ij}) = I_{g_j}(z_i)\, I_{f,p}(z_j)$, where z_i and z_j denote the i-th and the j-th bits in input to g_j and f, respectively.*

(b) *The average sensitivity of h is $s(h) = \sum_{j=1}^{m} (I_{f,p}(z_j)\, s(g_j))$.*

Proof.

(a) Follows from the definitions of influence and on-line influence.

(b) Follows from the definition of average sensitivity as sum of the influences of all the variables.

∎

Corollary 13

1. *$s(h) \leq \sum_{j=1}^{m} s(g_j)$.*
2. *$s(h) \leq \max_{1\leq j\leq m}\{s(g_j)\} s_p(f) \leq \max_{1\leq j\leq m}\{s(g_j)\} s_{max}(f)$.*
3. *If $g_j = g$ for all j, then $s(h) = s(g)\, s_p(f)$.*
4. *If $g_j \in \{\vee\, , \wedge\}$ for all j, then $s(h) \leq \frac{mk_{min}}{2^{k_{min}-1}} \leq \frac{n}{2^{k_{min}-1}}$, where $k_{min} = \min_{1\leq i\leq m} k_i$.*
5. *If ϕ is a Boolean function computable by a read-once formula and by a* layered *circuit of depth k, then $s(\phi) \leq s(f_1) \prod_{i=2}^{k} s_p(f_i)$, where the functions f_i satisfy*
   - *The output of f_k is the value computed by ϕ*
   - *For $i \neq k$, f_i is the function of maximal on-line sensitivity whose output is an input for f_{i+1}.*

Proof.

1. Follows from the fact that $I_{f,p}(z_j) \leq 1$ for $1 \leq j \leq m$..
2. Follows from

$$\begin{aligned} s(h) &= \sum_{j=1}^{m} I_{f,p}(z_j) s(g_j) \leq \max_{1\leq j\leq m}\{s(g_j)\} \sum_{j=1}^{m} I_{f,p}(z_j) = \max_{1\leq j\leq m}\{s(g_j)\}\ s_p(f) \\ &= \max_{1\leq j\leq m}\{s(g_j)\} \sum_{w} p_w s_f(w) \leq \max_{1\leq j\leq m}\{s(g_j)\} \max_{w}\{s_f(w)\}. \end{aligned}$$

3. Follows from

$$s(h) = \sum_{j=1}^{m} I_{f,p}(z_j) s(g_j) = \sum_{j=1}^{m} I_{f,p}(z_j) s(g) = s(g) s_p(f).$$

4. Follows from 2 and from the following facts:
   - $\max_{1\leq j\leq m}\{s(g_j)\} = \max_{1\leq j\leq m}\left\{\frac{k_j}{2^{k_j-1}}\right\} = \frac{k_{min}}{2^{k_{min}-1}}$, with $k_{min} = \min_{1\leq i\leq m}\{k_i\}$.
   - $s_{max}(f) \leq m$.
   - $mk_{min} \leq \sum_{j=1}^{m} k_j = n$.

   Note that for $k_{min} = 2$, we get $s(h) \leq \frac{n}{2}$, and for $k_{min} = \log n + 1$, $s(h) \leq 1$.

5. Follows by repeated applications of 2.

■

The above corollary has several interesting consequences, e.g. the fact that the average sensitivity of functions computable by read-once formulas is upper bounded by $\frac{n}{2}$, and that, if the minimum fan-in of a gate is of order $\log n$, then the average sensitivity is not increasing.

6 Applications to complexity

In this Section we present some applications of Fourier Analysis to Circuit Complexity. We first review a result by Linial et al. (see [LMN 93]) about a spectral characterization of AC^0 functions. We then find connections between *average sensitivity*, sparseness and complexity, and we use these relations to prove some results on symmetric AC^0 functions. We define some new complexity classes, essentially composed of functions that have most of their power spectrum in a relatively narrow contiguous band. We finally derive a version of Khrapchenko's theorem expressed in terms of Fourier coefficients.

6.1 Symmetric AC^0 functions

It is well known that constant depth circuits require exponential size to compute the parity function. More precisely AC^0-circuits cannot even approximate the parity function. This fact has some consequences on the Fourier transform, because, as showed is Section 2, the Fourier coefficients measure the correlation between a function and the parity of subsets of the input bits. Consequently, high-order Fourier coefficients of AC^0 functions must be very small. By exploiting this fact, Linial et al. (see [LMN 93]) were able to give the following spectral characterization of the class AC^0.

Lemma 14 ([LMN 93])
Let f be a Boolean function on n variables computable by a Boolean circuit of depth d and size M, and let t be any integer. Then

$$\sum_{|S|>t} (f^*(S))^2 \leq 2M \ 2^{\frac{-t^{1/d}}{20}} .$$

An application of this lemma implies:

Lemma 15 ([LMN 93])
For any function $f \in AC^0[d]$, we have $s(f) = O((\log n)^d)$.

This bound can be dramatically improved in the case of symmetric AC^0 functions. First of all, we show that there is a correspondence between sparsity or co-sparsity and functions whose average sensitivity decreases exponentially. This fact is the basis for showing that low average sensitivity is a structural property of sets which generalizes sparsity in a natural way. Then we find some interesting relations between a measure of complexity for symmetric functions defined in [FKPS 85] and the average sensitivity. These relationships allow us to use a result of [CK 91] for proving that the average sensitivity of symmetric functions - and, more in general, of functions with polynomial index - in AC^0 decreases exponentially.

Definition 12 *We say that a language over the alphabet* $\{0,1\}$ *is*

(i) generalized sparse *if, for any* n*, the number of strings of length* n *which belong to it is either at most* $n^{O(1)}$ *or at least* $2^n - n^{O(1)}$*;*

(ii) generalized almost sparse *if the number of strings of length* n *which belong to it is either at most* $n^{polylogn}$ *or at least* $2^n - n^{polylogn}$.

Note that if we have a generalized sparse language L (for which we know, for any length n, if it has at most $n^{O(1)}$ or at least $2^n - n^{O(1)}$ elements) then SAT cannot be reduced to L (assuming $P \neq NP$). Indeed there is a language L' that L (and therefore, SAT) is 1-truth-table reducible to, and by [OW 91], this implies $P = NP$.

Lemma 16 *A set is generalized sparse iff the average sensitivity of its characteristic function is* $O\left(\frac{n^{O(1)}}{2^n}\right)$*. A set is generalized almost sparse iff the average sensitivity of its characteristic function is* $O\left(\frac{n^{polylogn}}{2^n}\right)$.

Proof. From the definition of average sensitivity, we have that $s(f) \leq 2n\ min\{p,q\}$, where $p = Pr\{f(w) = 1\}$, $q = 1-p$, and the probability is taken over all $w \in \{0,1\}^n$. The *if* part of the Lemma follows from the fact that the characteristic function of a generalized sparse set satisfies either $p \leq \frac{n^{O(1)}}{2^n}$ or $q \leq \frac{n^{O(1)}}{2^n}$. It remains to prove that the functions with exponentially low average sensitivity are the characteristic functions of a generalized sparse set. From Parseval's identity [Leh 71] and from the definition of f_0^* we obtain $p = \sum_i (f_i^*)^2$ and $p = f_0^*$. Then, we get

$$p = \sum_i (f_i^*)^2 = f_0^{*2} + \sum_{i \neq 0} (f_i^*)^2 \leq f_0^{*2} + \sum_{i \neq 0} |i|(f_i^*)^2 = f_0^{*2} + \frac{s(f)}{4},$$

i.e. $p^2 - p + \frac{s(f)}{4} \geq 0$. Solving the latter inequality for p and using the hypothesis that $s(f) \leq \frac{n^{O(1)}}{2^n}$ we obtain that either $p \leq \frac{n^{O(1)}}{2^n}$ or $q \leq \frac{n^{O(1)}}{2^n}$, i.e. f is the characteristic function of a generalized sparse set. The proof for generalized almost sparse sets is similar.

■

Let f be a symmetric Boolean function. We recall some definitions from [FKPS 85]. The minimum number of variables of f that have to be set to constant values so that f becomes a constant function is called *measure* of f and is denoted by $\mu(f)$. Let $w \in \{0,1\}^{n+1}$ with elements w_i, where w_i is equal to the value of f when i variables are set to 1 and the other variables are set to 0. w is called *spectrum* of f. w_i is called i-th *character* of w. A *subword* of the spectrum is a string of the form $w_i\, w_{i+1} \ldots w_{i+k}$. [FKPS 85] show that $\mu(f)$ can be easily evaluated from the spectrum because $\mu(f) = n+1-\Gamma$, where Γ is the length of the longest constant subword of w.

Lemma 17 *Let f be a symmetric Boolean function defined on $\{0,1\}^n$ with measure $\mu(f)$. Then*

$$\frac{n}{2^{n-1}}\sum_{k=0}^{\bar{k}}\left\{\binom{n-1}{\lfloor\frac{\mu(f)}{2}\rfloor-k\Gamma-1}+\binom{n-1}{\lceil\frac{\mu(f)}{2}\rceil-k\Gamma-1}\right\}\le s(f)\le\frac{n}{2^{n-1}}\sum_{k=0}^{\mu(f)-1}\binom{n-1}{k},$$

where $\bar{k}=\lfloor\frac{\mu(f)}{2\Gamma}\rfloor$. Furthermore, both the lower and the upper bounds are tight.

Proof. The influence of x_i on a symmetric function f is given by

$$I_f(x_i) = \frac{1}{2^{n-1}}\sum_{k=0}^{n-1}\binom{n-1}{k}|w_{k+1}-w_k|,$$

where w_k denotes the k-th character of the spectrum of f. Hence we have

$$s(f) = \frac{n}{2^{n-1}}\sum_{k=0}^{n-1}\binom{n-1}{k}|w_{k+1}-w_k|.$$

We need evaluate the maximum value which can be attained by the average sensitivity for functions whose measure is $\mu(f)$. By analyzing the behavior of the binomial coefficients $\binom{n-1}{k}$ for $0\le k\le n-1$, it is easy to see that this maximum can be detected by looking at spectra of one of the forms

$$w=\underbrace{1010101\ldots10}_{\mu}\underbrace{11111\ldots1}_{\Gamma};$$

$$w=\underbrace{11111\ldots1}_{\Gamma}\underbrace{0101010\ldots10}_{\mu}$$

$$w=\underbrace{1010101\ldots01}_{\mu}\underbrace{00000\ldots0}_{\Gamma};$$

$$w=\underbrace{00000\ldots0}_{\Gamma}\underbrace{1010101\ldots01}_{\mu}$$

Thus, the maximum average sensitivity of functions with measure $\mu(f)$ is

$$\frac{n}{2^{n-1}}\sum_{k=0}^{\mu(f)-1}\binom{n-1}{k}.$$

The lower bound on $s(f)$ can be evaluated similarly. In fact, the functions of measure $\mu(f)$ whose average sensitivity is minimal must have spectra of the form:

$$\underbrace{000\ldots0}_{\lfloor\frac{\mu(f)}{2}\rfloor-\bar{k}\Gamma}\underbrace{\underbrace{111\ldots1}_{\Gamma}\underbrace{000\ldots0}_{\Gamma}\ldots\underbrace{111\ldots1}_{\Gamma}}_{\bar{k}\Gamma}\underbrace{000\ldots0}_{\Gamma}\underbrace{\underbrace{111\ldots1}_{\Gamma}\ldots\underbrace{000\ldots0}_{\Gamma}\underbrace{111\ldots1}_{\Gamma}}_{\bar{k}\Gamma}\underbrace{000\ldots0}_{\lceil\frac{\mu(f)}{2}\rceil-\bar{k}\Gamma},$$

where $\bar{k} = \lfloor \frac{\mu(f)}{2\Gamma} \rfloor$. Hence there exist functions of average sensitivity

$$\frac{n}{2^{n-1}} \sum_{k=0}^{\bar{k}} \left\{ \left(\begin{array}{c} n-1 \\ \lfloor \frac{\mu(f)}{2} \rfloor - k\Gamma - 1 \end{array} \right) + \left(\begin{array}{c} n-1 \\ \lceil \frac{\mu(f)}{2} \rceil - k\Gamma - 1 \end{array} \right) \right\},$$

which is the minimum possible value for functions of measure μ.

∎

The above Lemma has a very interesting consequence for symmetric functions in AC^0.

Theorem 18 *Let f be a symmetric function in AC^0. Then $s(f) = O(2^{-n+polylog\, n})$, and this is equivalent to saying that f is the characteristic function of a generalized almost sparse language.*

Proof. A consequence of lemma 17, together with the characterization given by [FKPS 85], is that symmetric functions in AC^0 have exponentially decreasing average sensitivity. In fact, since symmetric functions in AC^0 have measure bounded above by a polylog, from lemma 17 we obtain

$$s(f) \leq \frac{n}{2^{n-1}} \sum_{k=0}^{polylogn} \left(\begin{array}{c} n-1 \\ k \end{array} \right) \leq \frac{n^{polylogn}}{2^n}.$$

Then lemma 16 of the previous Section implies that f is the characteristic function of a generalized almost sparse language.

∎

Corollary 19 *$\theta\left(n^{polylogn}\right)$ symmetric functions of n variables are computable by polynomial size constant depth circuits.*

Proof. The upper bound follows from Theorem 1 in [WWY 92] and standard counting arguments. The lower bound is obtained by counting the number of functions for which $\mu(f) = O(polylog\, n)$.

∎

We also have the following result for symmetric AC^0 functions. Let $f : \{0,1\}^n \rightarrow \{0,1\}$ be a Boolean function and $p(x_1, \ldots, x_n)$ a real multivariate polynomial. Following [Pa 92], we say that a real multivariate polynomial *approximates* f if, for every $w \in \{0,1\}^n$, $|p(w) - f(w)| \leq \frac{1}{3}$. The *approximate degree* $d(f)$ of f is then defined to be the minimum, over all polynomials p that approximate f, of the degree of p. From [Pa 92] and [WWY 92], we easily obtain

Proposition 20 *Let $f \in AC^0$ be a non-constant symmetric function. Then, the approximate degree of f is $d(f) = \Theta\left(\sqrt{n \text{ polylog } n}\right)$.*

The previous results show that symmetric functions in AC^0 are such that the influence of variables must be almost exponentially small. The same is clearly not true for non symmetric functions. Thus if one wants to look for interesting functions in AC^0, she should concentrate on non-symmetric functions. As an example we analyze the Fourier spectrum of a family of $AC^0[2]$ functions for which the influence of variables is only polynomially small.

Let f be the following Boolean function $f(x_1, x_2, \ldots, x_n) = \bigwedge_{i=0}^{\frac{n}{k}-1} \left(\bigvee_{j=ik+1}^{(i+1)k} x_j \right)$.

Proposition 21 *For any $w \in \{0,1\}^n$, $|f^*(w)| = \frac{1}{2^n}(2^k - 1)^l$ where the integer l is such that*

$$\frac{n}{k} - |w| \ \le \ l \ \le \ \left\lfloor \frac{n - |w|}{k} \right\rfloor \qquad \text{for} \ \ |w| < \frac{n}{k},$$

and

$$0 \ \le \ l \ \le \ \left\lfloor \frac{n - |w|}{k} \right\rfloor \qquad \text{for} \ \ |w| \ge \frac{n}{k}.$$

Note that the maximum coefficient is $f^*(w) = \frac{1}{2^n}(2^k - 1)^{\frac{n}{k}}$ and it corresponds to the string w s.t. $|w| = 0$, while the minimum of the absolute value of the coefficients is $|f^*(w)| = \frac{1}{2^n}$.

Proposition 22 *The absolute value of the t-th Fourier coefficient f_t^* is given by $|f_t^*| = \frac{(2^k-1)^{\frac{n}{k}-h-\alpha}}{2^n}$, where $t = \sum_{s=1}^{h} i_s 2^{j_s k} + \alpha i_{h+1}$, and $0 \le h \le \frac{n}{k} - 1$; $\alpha \in \{0,1\}$; $i_s \in \{1, 2, \ldots, 2^k - 1\}$, $s = 1, 2, \ldots, h+1$; $j_s \in \{1, 2, \ldots, \frac{n}{k} - 1\}$, $s = 1, 2, \ldots, h$; $j_p \neq j_q$ $\forall p, q$.*

Note that the cardinality of the binary expansion of the integer t is given by $|t| = \sum_{s=1}^{h+1} |i_s|$, where $|i_s|$ is defined as the cardinality of the binary expansion of i_s.

6.2 New complexity classes

The symmetry in the average sensitivity distribution (see the proof of lemma 7) suggests the definition of complexity classes, which we call pAC^0 (for *parity* AC^0), and sAC^0 (for *symmetric* AC^0), where

$$\begin{aligned} pAC^0 &= \{g | g = f \oplus PARITY\ ,\ f \in AC^0\}\ , \\ sAC^0 &= AC^0 \cup pAC^0\ . \end{aligned}$$

These classes have the following properties.

- For any Boolean function $g \in pAC^0$ we have $s(g) = w(n - \log^k n)$, for a constant k, i.e. functions in pAC^0 behave similarly to $PARITY$ or its complement (like functions in AC^0 behave similarly to the two constant functions).

- sAC^0 is a class which lies slightly above AC^0. In fact $sAC^0 \subset ACC$.

- We can use Theorems 2.2 and 2.8 in [Leh 71], to prove that, if $g = f \oplus PARITY$, then
$$g^*(w) = \frac{1}{2}\delta_{w,0} - \frac{1}{2}\delta_{w \oplus u, 0} + f^*(w \oplus u)\ ,$$
where u is the vector whose entries are all equal to 1, and $\delta_{i,j}$ is the Kronecker delta function. Thus the Fourier coefficients of order $|i|$ of g, $1 \le |i| < n$, coincide with the Fourier coefficients of order $n - |i|$ of f, while $g_0^* = \frac{1}{2} + f_{2^n-1}^*$, and $g_{2^n-1}^* = -\frac{1}{2} + f_0^*$.

This last symmetric property of the Fourier coefficients allows us to adapt results on the Fourier coefficients of AC^0 functions to pAC^0 functions. As an example, we have that functions in pAC^0 have almost all their power spectrum on the high order coefficients.

In general, it is interesting to compare the spectrum of a function f with that of a function g defined as $g(w) = f(w) \oplus p_m(w)$, where p_m is the parity of m bits, e.g. the first m bits. From the fact that $p_m(w) = w^T a (mod\ 2)$, where $a^T = [1, 1, \ldots, 1, 0, \ldots, 0]$, some algebra and the application of Theorem 2.8 in [Leh 71] yield

$$g^*(w_1, \ldots, w_n) = f^*(\bar{w_1}, \ldots, \bar{w_m}, w_{m+1}, \ldots, w_n),$$

for $w \neq 0$ and $w \neq a$, and $g^*(0) = \frac{1}{2} + f^*(a)$, $g^*(a) = -\frac{1}{2} + f^*(0)$. Thus, all - but those for $w = 0$ and $w = a$ - the Fourier coefficients of f and g are the same, in different order. Some algebraic manipulation yields

$$s(g) = m - s(f) + 8 \sum_{w \neq 0, w \neq a} w^T (u - a)(f^*(w \oplus a))^2,$$

from which $s(g) \geq m - s(f)$.

The application of the above arguments to AC^0 functions gives rise to the definition of the complexity classes

$$\begin{aligned} p_i AC^0 &= \{g | g = f \oplus p_i, \ f \in AC^0\} \\ cAC^0 &= \cup_{i=0}^{n} p_i AC^0, \end{aligned}$$

where $p_0 AC^0 = AC^0$. We have

$$AC^0 \subset cAC^0 \subset ACC.$$

Generalizing further, we can define a class of Boolean functions that has at most k nonzero Fourier coefficients, or that has most of the power spectrum concentrated in such coefficients. It is not hard to see that every such class with $k = 2^{polylog}$ (as is the case for AC^0) is probably (approximately) learnable in time $2^{polylog}$ under uniform distribution. Simply sample the function at $2^{polylog}$ randomly chosen points, and determine the unknown Fourier coefficients solving the linear equations given by the matrix H.

6.3 Formula size lower bounds

Khrapchenko's theorem, which gives lower bounds on the size of Boolean formulas, can be restated in terms of Fourier coefficients as follows. Following [Kou 93], let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function on n variables, let $L(f)$ denote the number of leaves of the minimal-size $\{\vee, \wedge, \neg\}$-formula that computes f, and p the probability that f takes the value 1.

Lemma 23

$$L(f) \geq \frac{1}{4p(1-p)} s(f)^2 = \frac{4}{p(1-p)} \left(\sum_{w} |w| (f^*(w))^2 \right)^2.$$

Proof. Let $A = f^{-1}(0)$ and $B = f^{-1}(1)$. Following [Kou 93], we define the $|A| \times |B|$ matrix Q, with $q_{ij} = 1$ if $i \in B$ e $j \in A$ differ in exactly one variable; otherwise, $q_{ij} = 0$. Khrapchenko's theorem can be restated as follows: $L(f) \geq \frac{1}{|A| \cdot |B|} \left(\sum_{ij} q_{ij} \right)^2$. Then, the claim follows by noting that $\sum_{ij} q_{ij} = 2^{n-1} s(f) = 2^{n+1} \sum_w |w| (f^*(w))^2$, $|B| = 2^n p$ and $|A| = 2^n (1-p)$.

■

A number of corollaries follows from this Lemma, e.g. $L(f) \geq \max_w |w|^2 (f^*(w))^4$.

7 Concluding Remarks and New Research Directions

This paper has presented a complete framework in which to analyze the Fourier Transform of Boolean Functions. In addition to a review of known results, we have given an interpretation of the harmonic analysis on the hypercube in terms of its *Laplacian Matrix*, which has allowed us to describe in a natural way the interplay between Fourier coefficients and the *sensitivity* of Boolean functions. We have shown the relationship between the Fourier coefficients and the influences of the variables, and the connections between *average sensitivity*, sparseness and complexity, and used them to find properties of symmetric AC^0 functions. We have given a classification of functions according to their average sensitivity and introduced a natural generalization of the class AC^0. Finally, we have derived a version of Khrapchenko's theorem expressed in terms of Fourier coefficients.

We hope that the Fourier approach can be further exploited to study the complexity of Boolean functions.

References

[B93] R. Beigel. *The Polynomial Method in Circuit Complexity.* Proc. 8th Structure, pp. 82-95 (1993).

[BC 93] A. Bernasconi, B. Codenotti. *Sensitivity of Boolean Functions, Abstract Harmonic Analysis, and Circuit Complexity.* Technical Report TR-93-030 International Computer Science Institute, Berkeley, CA. (1993).

[BC 94] A. Bernasconi, B. Codenotti. *Measures of Boolean Function Complexity based on Harmonic Analysis.* Lecture Notes in Computer Science 778, pp. 63-72 (1994).

[BL 89] M. Ben Or, N. Linial. *Collective Coin Flipping.* In *Advances in Computing Research - Randomness and Computation* **5**, pp. 91-115, JAI Press (1989).

[Bop 89] R. B. Boppana. *Amplification of Probabilistic Boolean Formulas.* In *Advances in Computing Research - Randomness and Computation* **5**, pp. 27-46, JAI Press (1989).

[BOH 90] Y. Brandman, A. Orlitsky, J. Hennessy. *A Spectral Lower Bound Technique for the Size of Decision Trees and Two-Level AND/OR Circuits.* IEEE Trans. on Computers **39(2)**, pp. 282-287 (1990).

[Bru 90] J. Bruck. *Harmonic Analysis of Polynomial Threshold Functions.* SIAM Journal on Discrete Mathematics **3**, pp. 168-177 (1990).

[BS 92] J. Bruck, R. Smolensky. *Polynomial Threshold Functions, AC^0 Functions, and Spectral Norms.* SIAM Journal on Computing **21(1)**, pp. 33-42 (1992).

[BT 96] N.H. Bshouty, C. Tamon. *On the Fourier spectrum of monotone functions.* Journal of the ACM **43(4)** pp. 747-770 (1996).

[CK 91] P. Clote, E. Kranakis. *Boolean Functions, Invariance Groups, and Parallel Complexity.* SIAM Journal on Computing **20(3)**, pp. 553-590 (1991).

[DZ 97] M. Dubiner, U. Zwick. *Amplification and Percolation.* SIAM Journal on Computing **26(1)**, pp. 15-38 (1997).

[FKPS 85] R. Fagin, M.M. Klawe, N.J. Pippenger, L. Stockmeyer. *Bounded-Depth, Polynomial Size Circuits for Symmetric Functions.* Theoretical Computer Science **36**, pp. 239-250 (1985).

[GL 94] C. Gotsman, N. Linial. *Spectral Properties of threshold functions.* Combinatorica **14**, pp. 35-50 (1994).

[HMM 82] S.L. Hurst, D.M. Miller, J.C. Muzio. *Spectral Method of Boolean Function Complexity.* Electronics Letters **18(33)**, pp.572-574 (1982).

[Jac 95] J. Jackson. *The Harmonic Sieve: A Novel Application of Fourier Analysis to Machine Learning Theory and Practice.* PhD Thesis, Carnagie Mellon University (1995).

[KKL 88] J. Kahn, G. Kalai, N. Linial. *The Influence of Variables on Boolean Functions.* Proc. 29th FOCS, pp. 68-80 (1988).

[Kar 76] M.G. Karpovsky. *Finite Orthogonal Series in the Design of Digital Devices.* John Wiley and Son, New York (1976).

[Kou 93] E. Koutsoupias. *Improvements on Khrapchenko's Theorem.* Theoretical Computer Science **116**, pp. 399-403 (1993).

[KM 93] E. Kushilevitz, Y. Mansour. *Learning Decision Trees using the Fourier Spectrum.* SIAM Journal on Computing **22(6)** pp. 1331-1348 (1993).

[Leh 71] R. J. Lechner. *Harmonic Analysis of Switching Functions.* In *Recent Development in Switching Theory*, Academic Press, pp. 122-229 (1971).

[LMN 93] N. Linial, Y. Mansour, N. Nisan. *Constant Depth Circuits, Fourier Transform, and Learnability.* Journal of the ACM **40(3)**. pp. 607-620 (1993).

[Lit 40] D. Littlewood. *The Theory of Group Characters and the Matrix Representation of Groups.* Oxford University Press, London (1940).

[Loo 53] L.M. Loomis. *An Introduction to Abstract Harmonic Analysis.* Van Nostrand, Princeton, New Jersey (1953).

[Man 94] Y. Mansour. *Learning Boolean Functions via the Fourier Transform.* Theoretical Advances in Neural Computing and Learning, Kluwe Academic Publishers (1994).

[OW 91] M. Ogiwara, O. Watanabe. *On Polynomial-time Bounded Truth-table Reducibility of NP sets to Sparse sets.* SIAM Journal on Computing **20(3)**, pp. 471-483 (1991).

[SHB 68] F. Sellers, M.Y. Hsiao, L.W. Bearnson. *Error Detecting Logic for Digital Computers.* McGraw Hill (1968).

[Pa 92] R. Paturi. *On the Degree of Polynomials that Approximate Symmetric Boolean Functions.* Proc. 24th STOC, pp. 468-474 (1992).

[SV 85] S. Skyum, L.G. Valiant. *A Complexity Theory Based on Boolean Algebra.* Journal of the ACM **32(2)**, pp. 484-502 (1985).

[Sub 90] A. Subramanian. *The Computational Complexity of the Circuit Value and Network Stability Problems.* PhD Thesis, Tech. Report No. STAN-CS-90-1311, Stanford University (1990).

[WWY 92] I. Wegener, N. Wurm, S.Z. Yi. *Symmetric Functions in AC^0 can be Computed in Constant Depth with Very Small Size.* In "Boolean Function Complexity", Edited by M.S. Paterson, Cambridge University Press, pp. 129-139 (1992).