Electronic Colloquium on Computational Complexity, Report No. 174 (2012)

The Spectrum of Small DeMorgan Formulas

Anat Ganor* Ilan Komargodski* Ran Raz†

Abstract

We show a connection between the deMorgan formula size of a Boolean function and the noise stability of the function. Using this connection, we show that the Fourier spectrum of any balanced Boolean function computed by a deMorgan formula of size s is concentrated on coefficients of degree up to $O(\sqrt{s})$.

These results have several applications that apply to any function f that can be computed by a deMorgan formula of size s. First, we get that f can be approximated (in $\mathcal{L}_2$-norm) with constant error by a polynomial of degree $O(\sqrt{s})$. Second, we show an upper bound of $O(\sqrt{s})$ on the average sensitivity of f.

Our main result stems from a generalization of Khrapchenko's bound [Khr71], that might be of independent interest, and some Fourier analysis on the Boolean cube.

Previous works prove that any function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a deMorgan formula of size s, can be approximated point-wise by a polynomial of degree $O(s^{1/2+o(1)})$ with constant point-wise error. We note that this result can be easily extended to have a polynomial of degree $O(t \cdot s^{1/2+o(1)})$ that approximates f with point-wise error 2^{-t}, for any $t > 0$. This was shown in a long line of results in quantum complexity, including [BBC+01] and [FGG08, ACR+07, RS08, Rei09].

1 Introduction

In the seminal paper of Linial, Mansour and Nisan [LMN93], it is shown that every Boolean function that can be computed by an $\mathbf{AC^0}$ circuit, has a low-degree polynomial that approximates the function with error exponentially decreasing in the degree. Construction of *low-degree* polynomials that approximate Boolean functions is a central tool in complexity theory that has numerous applications. In particular, the result of [LMN93] has various applications in many fields such as learning theory, cryptography, pseudorandomness and derandomization.

*Weizmann Institute of Science, Rehovot 76100, Israel. Email:{anat.ganor,ilan.komargodski}@weizmann.ac.il.
Research supported by an Israel Science Foundation grant and by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation.

†Weizmann Institute of Science, Rehovot 76100, Israel and Institute for Advanced Study, Princeton, NJ. Email:ran.raz@weizmann.ac.il.
Research supported by an Israel Science Foundation grant and by the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation.

ISSN 1433-8092

In this work, we show several results regarding deMorgan formulas. A *deMorgan formula* is a Boolean formula over the basis $B_2 = \{\vee, \wedge, \neg\}$ with fan in at most 2. A deMorgan formula is represented by a tree such that every leaf is labeled by an input variable and every internal node is labeled by an operation from B_2. A formula is said to compute a function $f : \{0,1\}^n \to \{0,1\}$ if on input $x \in \{0,1\}^n$ it outputs $f(x)$. The computation is done in the natural way from the leaves to the root. The size of a formula is defined as the number of leaves it contains. For a Boolean function f we denote by $L(f)$ the size of the smallest deMorgan formula that computes f.

We show a connection between the deMorgan formula size of a Boolean function and the noise stability of the function. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. For every $p \in (0, 1/2]$ we show that

$$\mathbf{NS}_p(f) \geq 1 - 2p\sqrt{L(f) \cdot \|f\|^2 \cdot (1 - \|f\|^2)}$$

where $\|f\|$ denotes the $\mathcal{L}_2$-norm of f, and $\mathbf{NS}_p(f)$ is the noise stability of f with parameter p, as formally defined in Definition 2.9.

In addition, we show that the Fourier spectrum of a balanced Boolean function is concentrated on coefficients of degree up to $O(\sqrt{L(f)})$. More formally, for every Boolean function $f : \{0,1\}^n \to \{0,1\}$ and every $\varepsilon > 0$, we show that for $k = \sqrt{\frac{L(f)\cdot(1-\|f\|^2)}{\varepsilon^2 \cdot \|f\|^2}}$ it holds that

$$\sum_{\substack{S \subseteq [n], \\ |S| < k}} \widehat{f}(S)^2 \geq \|f\|^2 (1 - \varepsilon)$$

where $\widehat{f}(S)$ denotes the Fourier coefficient of f at S. This implies that f can be approximated (in $\mathcal{L}_2$-norm), with error $\varepsilon\|f\|^2$, by a polynomial of degree $< k$. Notice that if $\|f\|^2 < 1/2$ then one may prefer to approximate $1 - f$ rather than f. We note that the quadratic dependence between $L(f)$ and the degree of the approximating polynomial is tight, since the parity function over n variables is computed by a deMorgan formula of size $\Theta(n^2)$ and its Fourier representation is concentrated on the largest coefficient.

Another application of our results is an upper bound on the average sensitivity of f. We show that $\mathbf{AS}(f) \leq O(\sqrt{L(f)} \cdot \|f\|^2 \cdot (1 - \|f\|^2))$, where $\mathbf{AS}(f)$ denotes the average sensitivity of f, as formally defined in Definition 8.2.

1.1 Previous Work

Previous works give an upper bound on the degree of an approximating polynomial using tools from quantum complexity. Specifically, for every Boolean function f, Beals et al. [BBC+01] show that if f has a q-query bounded-error quantum algorithm (in the black box model), then there exists a polynomial of degree at most $2q$ that approximates f. Moreover, in a line of works in quantum query complexity [FGG08, ACR+07, RS08, Rei09] it is shown that if a Boolean function $f : \{0,1\}^n \to \{0,1\}$ can be computed by a deMorgan formula of size s, then there is a quantum black box algorithm that computes f in $O\left(\sqrt{s} \cdot {}^{\log n}/_{\log\log n}\right)$

queries, suffering from a point-wise error of $1/3$. By repeating independent applications of the algorithm, one can increase the number of queries to $O\left(t \cdot \sqrt{s} \cdot {\log n}/{\log\log n}\right)$ and reduce the point-wise error to 2^{-t}. Combining both of these results proves that every function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a deMorgan formula of size s can be approximated by a polynomial of degree $O\left(t \cdot \sqrt{s} \cdot {\log n}/{\log\log n}\right)$ up to point-wise error of 2^{-t}.

This result and our result are incomparable. Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function that can be computed by a deMorgan formula of size s. Our result gives $\mathcal{L}_2$-norm approximation which is *tight* in the degree (i.e, $O(\sqrt{s})$) for a constant ε. Moreover, our result is achieved using simple classical tools such as Khrapchenko's bound [Khr71] and Fourier analysis on the Boolean cube. The previous result is achieved using tools from quantum computing and quantum query complexity. The previous result gives a point-wise approximation by a polynomial which is almost optimal in the degree and with exponentially small point-wise error.

We note that there are results regarding the sign degree of functions that can be computed by small deMorgan formulas. The sign degree of a function is the minimal degree of a polynomial that agrees in *sign* with the function. In particular, combining the results of [FGG08, ACR$^+$07, RS08, Rei09] with the work of Lee [Lee09] fully resolves a conjecture by O'Donnell and Servedio [OS03] which states that the sign degree of every Boolean function that can be computed by a deMorgan formula of size s is $O(\sqrt{s})$.

As we have already mentioned, our main results (see Section 3) follow from a generalization of Khrapchenko's bound on the size of deMorgan formulas. Various generalizations of Khrapchenko's bound were used in the past in numerous works. Zwick [Zwi91] extended the definition of formula size to handle weighted input variables and generalized Khrapchenko's bound to cover the new definition. Koutsoupias [Kou93] was able to extend Khrapchenko's bound with a spectral version to give better lower bounds for specific functions. Håstad [Hås98] showed that the shrinkage exponent of Boolean deMorgan formulas (for the exact definition see [Hås98]) is 2. One of the components in his proof is a lower bound on the deMorgan formula size that depends on the probability that some restrictions occur (for the exact formulation see [Hås98]). Håstad proves that indeed this lower bound is a generalization of Khrapchenko's bound. Laplante, Lee and Szegedy [LLS06] introduce a new complexity measure for Boolean functions that is a lower bound on the deMorgan formula size. They show that several deMorgan formula size lower bounds (including [Khr71, Kou93, Hås98]) are, in fact, a special case of their method.

2 Preliminaries

We start with some general notation. We denote by $[n]$ the set of numbers $\{1, 2, \ldots, n\}$. For $i \in [n]$ and for $x \in \{0,1\}^n$, denote by x_i the i-th bit of x. We denote by $wt(x)$ the Hamming weight of a string $x \in \{0,1\}^n$ (i.e. the number of 1's in the string). We denote by $\Delta(x, y)$ the Hamming distance between two strings $x, y \in \{0,1\}^n$ (i.e. the number of coordinates in which x and y differ). In addition, for simplicity, we define $\frac{0}{0} = 0$.

2.1 DeMorgan Formulas

Throughout the paper we will only consider deMorgan formulas and not always explicitly mention it.

Definition 2.1. *A deMorgan formula is a Boolean formula with AND, OR and NOT gates with fan in at most 2.*

Definition 2.2. *The size of a formula F is the number of leaves in it and is denoted by $L(F)$. For a function $f : \{0,1\}^n \to \{0,1\}$, we will denote by $L(f)$ the size of the smallest formula computing the function f.*

2.2 Fourier Analysis

For each $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{-1,1\}$ as $\chi_S(x) = \prod_{i \in S}(-1)^{x_i}$. It is well known that the set $\{\chi_S\}_{S \subseteq [n]}$ is an orthonormal basis (called the Fourier basis) for the space of all functions $f : \{0,1\}^n \to \mathbb{R}$. It follows that every function $f : \{0,1\}^n \to \mathbb{R}$ can be represented as

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S)\chi_S(x)$$

where $\widehat{f} : \{0,1\}^n \to \mathbb{R}$, and $\widehat{f}(S)$ is called the Fourier coefficient of f at $S \subseteq [n]$.

Definition 2.3. *We define the inner product $\langle \cdot, \cdot \rangle$ on pairs of functions $f, g : \{0,1\}^n \to \mathbb{R}$ by*

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)g(x) = \mathop{\mathbb{E}}_{x \in \{0,1\}^n}[f(x)g(x)]$$

2.2.1 Basic Properties

Proposition 2.4. *For $f : \{0,1\}^n \to \mathbb{R}$ and $S \subseteq [n]$, the Fourier coefficient of f at S is*

$$\widehat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x)$$

Proposition 2.5. *Consider functions $f, g : \{0,1\}^n \to \mathbb{R}$. Since $\{\chi_S\}_{S \subseteq [n]}$ forms an orthonormal basis for the space of functions from $\{0,1\}^n$ to $\mathbb{R}$, we get Plancherel's theorem*

$$\langle f, g \rangle = \sum_{S,T \subseteq [n]} \widehat{f}(S)\widehat{g}(T) \langle \chi_S, \chi_T \rangle = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S)$$

In particular, the orthonormality of the basis gives the known Parseval theorem

$$\sum_{S \subseteq [n]} \left(\widehat{f}(S)\right)^2 = \langle f, f \rangle = \|f\|^2$$

where $\|f\|$ denotes the $\mathcal{L}_2$-norm of f.

2.2.2 Convolution

We begin by defining the convolution operation.

Definition 2.6 (Convolution)**.** *Let $f, g : \{0,1\}^n \to \mathbb{R}$. The convolution $f * g : \{0,1\}^n \to \mathbb{R}$ is defined as follows*

$$(f * g)(x) = \frac{1}{2^n} \sum_{y \in \{0,1\}^n} f(x \oplus y) g(y)$$

We state the well known convolution theorem.

Proposition 2.7 (The Convolution Theorem)**.** *Let $f, g : \{0,1\}^n \to \mathbb{R}$.*

$$\widehat{f * g}(S) = \widehat{f}(S)\widehat{g}(S)$$

2.3 Fourier Coefficients of Product Functions

We prove a simple lemma regarding Fourier coefficients of functions which are product functions. This lemma is useful to analyze Fourier coefficients of some specific functions.

Proposition 2.8. *Let $f : \{0,1\}^n \to \mathbb{R}$ be a function such that $f(x) = g(x_1) \cdot \dots \cdot g(x_n)$ for some function $g : \{0,1\} \to \mathbb{R}$. It holds that for $S \subseteq [n]$,*

$$\widehat{f}(S) = \widehat{g}(S_1) \cdot \dots \cdot \widehat{g}(S_n)$$

where $S_i = \{1\}$ if $i \in S$ and $S_i = \emptyset$ otherwise.

Proof. By the definition of Fourier coefficient (Proposition 2.4), we get that

$$\begin{aligned}
\widehat{f}(S) &= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x)\chi_S(x) \\
&= \frac{1}{2^n} \sum_{x \in \{0,1\}^n} g(x_1) \cdot \dots \cdot g(x_n) \chi_{S_1}(x_1) \dots \chi_{S_n}(x_n) \\
&= \left[\frac{1}{2} \sum_{x_1 \in \{0,1\}} g(x_1)\chi_{S_1}(x_1) \right] \cdot \dots \cdot \left[\frac{1}{2} \sum_{x_n \in \{0,1\}} g(x_n)\chi_{S_n}(x_n) \right] \\
&= \widehat{g}(S_1) \cdot \dots \cdot \widehat{g}(S_n)
\end{aligned}$$

as needed. □

2.4 Noise Stability

We define the *noise stability* of a Boolean function.

Definition 2.9 (Noise stability)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. For $p \in [0,1]$ and $x \in \{0,1\}^n$, define $N_p(x)$ to be the distribution of a random element $y \in \{0,1\}^n$ which satisfies $\Pr[x_i \neq y_i] = p$, independently for all $i \in [n]$. The p-noise stability of f is*

$$\mathbf{NS}_p(f) = \Pr_{\substack{x \in \{0,1\}^n, \\ y \sim N_p(x)}} [f(x) = f(y)]$$

3 Main Results

Let $f : \{0,1\}^n \to \{0,1\}$ be a function that can be computed by a small deMorgan formula and let $g : \{0,1\}^n \to \{0,1\}$ be such that $g(x) \leq f(x)$ for every $x \in \{0,1\}^n$. Our first theorem gives a lower bound on the noise stability of g.

Theorem 3.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a deMorgan formula of size s. Let $g : \{0,1\}^n \to \{0,1\}$ be a Boolean function such that $g^{-1}(1) \subseteq f^{-1}(1)$. Denote $\alpha = \frac{\left|g^{-1}(1)\right|}{2^n}$ and $\gamma = \frac{\left|f^{-1}(1) \setminus g^{-1}(1)\right|}{2^n}$. For any $p \in (0, 1/2]$, it holds that*

$$\mathbf{NS}_p(g) \geq 1 - 2\gamma - 2p\sqrt{s \cdot \alpha \cdot (1 - \alpha - \gamma)}$$

A useful corollary stating a lower bound on the noise stability of a function $f : \{0,1\}^n \to \{0,1\}$ that can be computed by a small deMorgan formula. This corollary stems from the previous theorem when setting $g^{-1}(1) = f^{-1}(1)$.

Corollary 3.2. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a deMorgan formula of size s. Denote $\alpha = \frac{|f^{-1}(1)|}{2^n}$. For any $p \in (0, 1/2]$, it holds that*

$$\mathbf{NS}_p(f) \geq 1 - 2p\sqrt{s \cdot \alpha \cdot (1 - \alpha)}$$

In addition, we show a lower bound on the Fourier weight of the "light" coefficients of f.

Theorem 3.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a deMorgan formula of size s. Denote $\alpha = \frac{|f^{-1}(1)|}{2^n}$. Then, for any $\varepsilon > 0$, letting $k = \frac{1}{\varepsilon}\sqrt{s\frac{1-\alpha}{\alpha}}$ it holds that*

$$\sum_{\substack{S \subseteq [n], \\ |S| < k}} \left(\widehat{f}(S)\right)^2 \geq \alpha\,(1 - \varepsilon)$$

4 Generalization of Khrapchenko's Bound

In this section we generalize the Khrapchenko bound on the size of deMorgan formulas. We begin by recalling the original Khrapchenko bound.

Theorem 4.1 ([Khr71])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and let $A \subseteq f^{-1}(1), B \subseteq f^{-1}(0)$. Denote $H(A,B) = \{(a,b) | a \in A, b \in B, \Delta(a,b) = 1\}$. It holds that*

$$L(f) \geq \mathcal{K}(A,B) = \frac{|H(A,B)|^2}{|A| \cdot |B|}$$

In this section we prove a lower bound for formula size that can be interpreted as a generalized version of Khrapchenko's theorem. Let $A, B \subseteq \{0,1\}^n$ and $p \in [0,1]$, we define

$$H_p(A,B) = \sum_{a \in A, b \in B} p^{\Delta(a,b)} (1-p)^{n-\Delta(a,b)}$$

Theorem 4.2 (Generalized Khrapchenko bound)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and let $A \subseteq f^{-1}(1), B \subseteq f^{-1}(0)$. It holds that for any $0 < p \leq 1$,*

$$L(f) \geq \mathcal{K}_p(A,B) = \frac{(H_p(A,B))^2}{|A| \cdot |B| \cdot p^2}$$

Proof. The proof follows the lines of the proof of Khrapchenko's bound from [Weg87] (see Section 8.8 there).

For $f : \{0,1\}^n \to \{0,1\}$ denote $K_p(f) = \max_{A \subseteq f^{-1}(1), B \subseteq f^{-1}(0)} \{\mathcal{K}_p(A,B)\}$. It is enough to prove that $K_p(f)$ is a formal complexity measure (see Lemma 8.1 in [Weg87]). In order to do so, we prove 3 properties of $K_p(f)$, following the original proof, as follows:

- $\forall i \in [n] : K_p(x_i) \leq 1$. Each vector in A (or B, symmetrically) contributes at most $\sum_{i=0}^{n-1} \binom{n-1}{i} p^{i+1}(1-p)^{n-i-1} = p$ to $H_p(A,B)$. It follows that $K_p(x_i) \leq 1$.
- $K_p(\neg f) = K_p(f)$. The definition of $K_p(f)$ is symmetric with respect to A and B.
- $K_p(f \vee g) \leq K_p(f) + K_p(g)$. We choose $A \subseteq (f \vee g)^{-1}(1)$ and $B \subseteq (f \vee g)^{-1}(0)$ such that $\mathcal{K}_p(A,B) = K_p(f \vee g)$. Since $B \subseteq (f \vee g)^{-1}(0)$, then $B \subseteq f^{-1}(0)$ and $B \subseteq g^{-1}(0)$. Partition A into disjoint $A_f \subseteq f^{-1}(1)$ and $A_g \subseteq g^{-1}(1)$. Then $H_p(A,B) = H_p(A_f,B) + H_p(A_g,B)$. Then,

$$K_p(f \vee g) = \frac{(H_p(A_f,B) + H_p(A_g,B))^2}{(|A_f| + |A_g|)|B|p^2}$$

$$K_p(f) + K_p(g) \geq \frac{(H_p(A_f,B))^2}{|A_f||B|p^2} + \frac{(H_p(A_g,B))^2}{|A_g||B|p^2}$$

The claim now follows (as done in [Zwi91]) since for every $a_1, a_2 \in \mathbb{R}$ and every $b_1, b_2 > 0$ it holds that

$$\frac{a_1^2}{b_1} + \frac{a_2^2}{b_2} \geq \frac{(a_1+a_2)^2}{b_1+b_2}$$

□

Remark: We consider our bound as a generalization of Khrapchanko's bound since for every Boolean function $f : \{0,1\}^n \to \{0,1\}$ and $A \subseteq f^{-1}(1), B \subseteq f^{-1}(0)$ it holds that

$$\lim_{p\to 0} \mathcal{K}_p(A,B) = \mathcal{K}(A,B)$$

We end this section with a lemma that will be useful for the rest of the paper.

Lemma 4.3. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function such that $L(f) = s$. For $A \subseteq f^{-1}(1)$, $B \subseteq f^{-1}(0)$ and $C = \{0,1\}^n \setminus (A \cup B)$ it holds that*

$$H_p(A,A) = |A| - H_p(A,B) - H_p(A,C)$$

and thus

$$H_p(A,A) \;\; \geq \;\; |A| - \sqrt{s \cdot |A| \cdot |B|} \cdot p - H_p(A,C)$$

Notice that when $C = \emptyset$ (that is $A = f^{-1}(1)$ and $B = f^{-1}(0)$), it holds that $H_p(A,C) = 0$.

Proof of Lemma 4.3. First, it is clear, by the definition, that $H_p(A,A) = H_p(A, A\cup B\cup C) - H_p(A,B) - H_p(A,C)$. Second, we notice that $H_p(A, A\cup B\cup C) = |A| \sum_{i=0}^{n} \binom{n}{i} p^i (1-p)^{n-i} = |A|$, which proves the equality of the lemma. For the second part, using Theorem 4.2 we get that $s = L(f) \geq \frac{(H_p(A,B))^2}{|A|\cdot|B|\cdot p^2}$. So $\sqrt{s \cdot |A| \cdot |B|} \cdot p \geq H_p(A,B)$ which proves the inequality of the lemma. □

4.1 Generalized Khrapchenko and Noise Stability

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function, let $A \subseteq f^{-1}(1)$, $B \subseteq f^{-1}(0)$ and $C = \{0,1\}^n \setminus (A \cup B)$. In this subsection we bound $\mathbf{NS}_p(f)$ in terms of $H_p(A,B)$ and $H_p(A,C)$.

Lemma 4.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function and $p \in [0,1]$. Let $A \subseteq f^{-1}(1)$, $B \subseteq f^{-1}(0)$ and $C = \{0,1\}^n \setminus (A \cup B)$. It holds that*

$$\mathbf{NS}_p(f) \geq \left(1 - \frac{|C|}{2^n}\right) - \frac{2H_p(A,B) + H_p(A\cup B, C)}{2^n}$$

Specifically, if $C = \emptyset$, it holds that

$$\mathbf{NS}_p(f) = 1 - \frac{2H_p(A,B)}{2^n}$$

Proof. By the definition of noise stability

$$\begin{aligned} \mathbf{NS}_p(f) &= \Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}} [f(x) = f(y)] \\ &\geq \Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}} [x \in A \wedge y \in A] + \Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}} [x \in B \wedge y \in B] \qquad (4.1) \end{aligned}$$

Using simple manipulations we get that

$$\begin{aligned}
\Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}}[x\in A\wedge y\in A] &= \sum_{x'\in\{0,1\}^n}\Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}}[x\in A\wedge y\in A|x=x']\Pr_{x\in\{0,1\}^n}[x=x'] \\
&= \frac{1}{2^n}\sum_{x'\in\{0,1\}^n}\Pr_{y\sim N_p(x)}[x\in A\wedge y\in A|x=x'] \\
&= \frac{1}{2^n}\sum_{x'\in A}\Pr_{y\sim N_p(x')}[y\in A] \\
&= \frac{1}{2^n}\sum_{x'\in A}\sum_{y'\in\{0,1\}^n}\Pr_{y\sim N_p(x')}[y\in A|y=y']\Pr_{y\sim N_p(x')}[y=y'] \\
&= \frac{1}{2^n}\sum_{x'\in A}\sum_{y'\in A}\Pr_{y\sim N_p(x')}[y=y'] \\
&= \frac{1}{2^n}\sum_{x'\in A}\sum_{y'\in A}p^{\Delta(x',y')}(1-p)^{n-\Delta(x',y')} \\
&= \frac{1}{2^n}H_p(A,A)
\end{aligned}$$

An analogous calculation shows that

$$\Pr_{\substack{x\in\{0,1\}^n,\\ y\sim N_p(x)}}[x\in B\wedge y\in B] = \frac{1}{2^n}H_p(B,B)$$

Plugging these back into equation (4.1), we get that

$$\begin{aligned}
\mathbf{NS}_p(f) &\geq \frac{1}{2^n}\left(H_p(A,A)+H_p(B,B)\right) \\
&= \frac{1}{2^n}\left(|A|+|B|-2H_p(A,B)-H_p(A\cup B,C)\right) \\
&= \left(1-\frac{|C|}{2^n}\right)-\frac{2H_p(A,B)+H_p(A\cup B,C)}{2^n}
\end{aligned}$$

where the first equality follows from Lemma 4.3.

Notice that if $C=\emptyset$, then the inequality in equation (4.1) becomes an equality (from which the equality in the lemma follows). □

5 Proof of Theorem 3.1

Let $f:\{0,1\}^n\to\{0,1\}$ be a Boolean function computable by a deMorgan formula of size s. Let $g:\{0,1\}^n\to\{0,1\}$ be a Boolean function such that $g^{-1}(1)\subseteq f^{-1}(1)$. Denote

$A = g^{-1}(1)$, $B = f^{-1}(0)$ and $C = f^{-1}(1) \setminus g^{-1}(1)$. Recall that $\alpha = \frac{|A|}{2^n}$ and $\gamma = \frac{|C|}{2^n}$. Notice that $\frac{H_p(A \cup B, C)}{2^n} \leq \gamma$ (by the definition of $H_p(A \cup B, C)$). Using Lemma 4.4 (applied for the function g) and Theorem 4.2, we get that

$$\begin{aligned}
\mathbf{NS}_p(g) &\geq (1-\gamma) - \frac{2H_p(A,B)}{2^n} - \gamma \\
&\geq 1 - 2\gamma - \frac{2 \cdot p\sqrt{s \cdot |A| \cdot |B|}}{2^n} \\
&= 1 - 2\gamma - 2p\sqrt{s \cdot \alpha \cdot (1-\alpha-\gamma)}
\end{aligned}$$

which proves Theorem 3.1.

6 Noise Stability and Fourier Expansion

In this section we prove a known relation between the noise stability of a Boolean function $f : \{0,1\}^n \to \{0,1\}$ and its Fourier expansion (see e.g. [BKS98, BJT99, O'D02]). We note that our analysis is similar to the analysis in previous proofs of this lemma (see the remark at the end of this section).

Lemma 6.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. For every $p \in (0, 1/2]$ it holds that*

$$\mathbf{NS}_p(f) = 1 - 2\widehat{f}(\emptyset) + 2\sum_{S \subseteq [n]} (1-2p)^{|S|} \left(\widehat{f}(S)\right)^2$$

Let $A, B \subseteq \{0,1\}^n$. Denote by I_A and I_B the characteristic functions of the sets A and B, respectively. In other words,

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise} \end{cases}, \quad I_B(x) = \begin{cases} 1 & x \in B \\ 0 & \text{otherwise} \end{cases}$$

Fix $p \in (0, 1/2]$ and denote by $\mathcal{I}_p : \{0,1\}^n \to [0,1]$ the function $\mathcal{I}_p(x) = p^{wt(x)}(1-p)^{n-wt(x)}$.

Lemma 6.1 is proved using Lemma 4.4 and the following lemma.

Lemma 6.2. *Let $f : \{0,1\}^n \to \{0,1\}$. For any $A, B \subseteq \{0,1\}^n$,*

$$H_p(A,B) = \sum_{a \in A, b \in B} p^{\Delta(a,b)}(1-p)^{n-\Delta(a,b)} = 2^{2n} \cdot \sum_{S \subseteq [n]} \widehat{I_A}(S)\widehat{I_B}(S)\widehat{\mathcal{I}_p}(S)$$

Proof. We can rewrite $H_p(A,B)$ as

$$\begin{aligned}
H_p(A,B) &= \sum_{x \in \{0,1\}^n} I_B(x) \sum_{y \in \{0,1\}^n} I_A(y)\mathcal{I}_p(x \oplus y) \\
&= 2^{2n} \cdot \langle I_A * \mathcal{I}_p, I_B \rangle
\end{aligned}$$

Expanding each function by the appropriate Fourier expansion and using the convolution theorem (Proposition 2.7) we get

$$\begin{aligned} H_p(A,B) &= 2^{2n} \cdot \left\langle \sum_{S\subseteq[n]} \widehat{I_A}(S)\widehat{\mathcal{I}_p}(S)\chi_S, \sum_{T\subseteq[n]} \widehat{I_B}(T)\chi_T \right\rangle \\ &= 2^{2n} \cdot \sum_{S\subseteq[n]} \widehat{I_A}(S)\widehat{I_B}(S)\widehat{\mathcal{I}_p}(S) \end{aligned}$$

which proves the claim. □

We will now investigate the Fourier coefficients of $\mathcal{I}_p$ and prove the following lemma.

Lemma 6.3. *For every $S \subseteq [n]$, it holds that*

$$\widehat{\mathcal{I}_p}(S) = \frac{(1-2p)^{|S|}}{2^n}$$

Proof. We first notice that $\mathcal{I}_p$ is a product distribution. That is, if we denote

$$I_p(x) = \begin{cases} p & x = 1 \\ 1-p & x = 0 \end{cases}$$

the function $I_p : \{0,1\} \to \mathbb{R}$ then

$$\mathcal{I}_p(x) = I_p(x_1) \cdot I_p(x_2) \cdot \dots \cdot I_p(x_n)$$

Using Proposition 2.8, it follows that for every $S \subseteq [n]$,

$$\widehat{\mathcal{I}_p}(S) = \prod_{i=1}^{n} \left(\widehat{I_p}(S_i)\right) \tag{6.1}$$

To calculate this, we prove a simple lemma about the Fourier coefficients of I_p.

Lemma 6.4. *It holds that $\widehat{I_p}(\{1\}) = \frac{1-2p}{2}$ and $\widehat{I_p}(\emptyset) = \frac{1}{2}$.*

Proof. By the definition, $\widehat{I_p}(\emptyset) = \mathbb{E}_{x\in\{0,1\}}[I_p(x)] = \frac{1}{2}$. The second part also follows directly from the definition of Fourier coefficients (Proposition 2.4). □

Plugging this lemma back into equation (6.1), we get that

$$\widehat{\mathcal{I}_p}(S) = \frac{(1-2p)^{|S|}}{2^n}$$

which completes the claim. □

We are now ready to prove the main lemma of this section.

Proof of Lemma 6.1. Let $f : \{0,1\}^n \to \{0,1\}$, let $A = f^{-1}(1)$, $B = f^{-1}(0)$. Fix $p \in [0,1]$. By Lemma 4.4 we know that

$$\mathbf{NS}_p(f) = 1 - \frac{2H_p(A,B)}{2^n} \tag{6.2}$$

By Lemma 4.3 we get that

$$H_p(A,B) = |A| - H_p(A,A)$$

Plugging Lemma 6.3 into Lemma 6.2 we get that

$$H_p(A,A) = 2^n \sum_{S \subseteq [n]} \left(\widehat{I_A}(S)\right)^2 (1-2p)^{|S|}$$

Denote $\alpha = \frac{|A|}{2^n}$. Plugging these into equation (6.2), we get that

$$\mathbf{NS}_p(f) = 1 - 2\alpha + 2 \sum_{S \subseteq [n]} \left(\widehat{f}(S)\right)^2 (1-2p)^{|S|}$$

which proves the lemma (recall that $\alpha = \widehat{f}(\emptyset)$). □

Remark: We note that our analysis is similar to the one done in the previous proofs of Lemma 6.1. The difference is that most of our analysis works for any $A \subseteq f^{-1}(1)$ and $B \subseteq f^{-1}(0)$ rather than $A = f^{-1}(1)$ and $B = f^{-1}(0)$ and might be of independent interest.

7 Proof of Theorem 3.3

Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a deMorgan formula of size s and let $\varepsilon > 0$. Let $A = f^{-1}(1)$, $B = f^{-1}(0)$. Denote $\alpha = \frac{|A|}{2^n}$ and $\beta = \frac{|B|}{2^n}$. Let $p \in (0, 1/2]$ (to be fixed later).

Using Theorem 3.1 and Lemma 6.1 we get that

$$1 - 2\alpha + 2 \sum_{S \subseteq [n]} (1-2p)^{|S|} \left(\widehat{f}(S)\right)^2 \geq 1 - 2p\sqrt{s \cdot \alpha \cdot \beta}$$

Splitting the sum of the left-hand side to "light" Fourier coefficients and "heavy" ones, we get that

$$\begin{aligned}
\alpha - p\sqrt{s \cdot \alpha \cdot \beta} &\leq \sum_{\substack{S \subseteq [n], \\ |S| < k}} \left(\widehat{f}(S)\right)^2 (1-2p)^{|S|} + \sum_{\substack{S \subseteq [n], \\ |S| \geq k}} \left(\widehat{f}(S)\right)^2 (1-2p)^{|S|} \\
&\leq \sum_{\substack{S \subseteq [n], \\ |S| < k}} \left(\widehat{f}(S)\right)^2 + (1-2p)^k \sum_{\substack{S \subseteq [n], \\ |S| \geq k}} \left(\widehat{f}(S)\right)^2
\end{aligned}$$

which means that for any $k \in [n]$,

$$\begin{aligned} \sum_{\substack{S\subseteq[n],\\|S|<k}} \left(\widehat{f}(S)\right)^2 &\geq \alpha - p\sqrt{s\cdot\alpha\cdot\beta} - (1-2p)^k \sum_{\substack{S\subseteq[n],\\|S|\geq k}} \left(\widehat{f}(S)\right)^2 \\ &= \alpha - p\sqrt{s\cdot\alpha\cdot\beta} - (1-2p)^k \left(\alpha - \sum_{\substack{S\subseteq[n],\\|S|<k}} \left(\widehat{f}(S)\right)^2\right) \end{aligned}$$

where the equality holds by Parseval's theorem (see Proposition 2.5). Finally, we get that

$$\begin{aligned} \sum_{\substack{S\subseteq[n],\\|S|<k}} \left(\widehat{f}(S)\right)^2 &\geq \alpha\left(\frac{1-p\sqrt{s\frac{\beta}{\alpha}}-(1-2p)^k}{1-(1-2p)^k}\right) \\ &= \alpha\left(1-\frac{p\sqrt{s\frac{\beta}{\alpha}}}{1-(1-2p)^k}\right) \end{aligned}$$

Setting $p = \frac{1}{2k}$, we get that

$$\begin{aligned} \sum_{\substack{S\subseteq[n],\\|S|<k}} \left(\widehat{f}(S)\right)^2 &\geq \alpha\left(1-\frac{\sqrt{s\frac{\beta}{\alpha}}}{2k\left(1-\left(1-\frac{1}{k}\right)^k\right)}\right) \\ &\geq \alpha\left(1-\frac{\sqrt{s\frac{\beta}{\alpha}}}{2k\left(1-e^{-1}\right)}\right) \\ &\geq \alpha\left(1-\frac{\sqrt{s\frac{\beta}{\alpha}}}{k}\right) \end{aligned}$$

Plugging in $k = \frac{1}{\varepsilon}\sqrt{s\frac{1-\alpha}{\alpha}}$ we get that

$$\sum_{\substack{S\subseteq[n],\\|S|<k}} \left(\widehat{f}(S)\right)^2 \geq \alpha\left(1-\varepsilon\right)$$

as needed.

8 Applications

In this section we survey some applications of our main theorems.

8.1 Approximating Formulas Using Low-Degree Polynomials

A useful consequence of Theorem 3.3 is that it is possible to approximate Boolean functions that are computed by small deMorgan formulas by low-degree polynomials. This fact is formally stated in the next corollary.

Corollary 8.1. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function computable by a deMorgan formula of size s. For every $\varepsilon > 0$, there exists a polynomial $p : \{0,1\}^n \to \mathbb{R}$ of degree $< \frac{1}{\varepsilon}\sqrt{s\frac{1-\alpha}{\alpha}}$ such that $\|f-p\|^2 \leq \varepsilon \cdot \alpha$, where $\alpha = \|f\|^2$.*

Proof. Let $k = \frac{1}{\varepsilon}\sqrt{s\frac{1-\alpha}{\alpha}}$ and let the polynomial $p(x)$ be defined as $p(x) = \sum_{\substack{S\subseteq[n],\\ |S|<k}} \widehat{f}(S)\chi_S(x)$.

Using Theorem 3.3, it follows that

$$\|f-p\|^2 \leq \varepsilon \cdot \alpha$$

as required. □

8.2 Average Sensitivity of Formulas

A consequence of Corollary 3.2 is a bound on the average sensitivity of functions that can be computed by small deMorgan formulas.

Definition 8.2 (Average sensitivity)**.** *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function. The average sensitivity (also known as total influence) of f is*

$$\mathbf{AS}(f) = \sum_{i=1}^{n} \Pr_{x\in\{0,1\}^n}\left[f(x) \neq f(x^{(i)})\right]$$

where $x^{(i)}$ is x with the i^{th} bit flipped.

Fact 8.3 ([KKL88])**.** $\mathbf{AS}(f) = 4\sum_{S\subseteq[n]} |S|\widehat{f}(S)^2$.

Corollary 8.4. *Let $f : \{0,1\}^n \to \{0,1\}$ be a Boolean function that can be computed by a deMorgan formula of size s and let $\alpha = \|f\|^2$. Then,*

$$\mathbf{AS}(f) \leq 2\sqrt{s\cdot\alpha\cdot(1-\alpha)}$$

Proof. Using Corollary 3.2 and Lemma 6.1, it follows that for any $p \in (0, 1/2]$,

$$\begin{aligned}
2p\sqrt{s\cdot\alpha\cdot(1-\alpha)} &\geq 2\widehat{f}(\emptyset) - 2\sum_{S\subseteq[n]}(1-2p)^{|S|}\left(\widehat{f}(S)\right)^2 \\
&\geq 2\widehat{f}(\emptyset) - 2\sum_{S\subseteq[n]}\left(\widehat{f}(S)\right)^2\left(1-2p|S| + \binom{|S|}{2}(2p)^2\right) \\
&= -2\sum_{S\subseteq[n]}\left(\widehat{f}(S)\right)^2\left(-2p|S| + \binom{|S|}{2}(2p)^2\right)
\end{aligned}$$

where the last equality holds since $\widehat{f}(\emptyset) = \sum_{S\subseteq[n]} \left(\widehat{f}(S)\right)^2$. Dividing both sides by p, we get that

$$2\sqrt{s\cdot\alpha\cdot(1-\alpha)} \;\;\geq\;\; -2\sum_{S\subseteq[n]} \left(\widehat{f}(S)\right)^2 \left(-2|S| + \binom{|S|}{2}4p\right)$$

Taking the limit when $p \to 0$, we get that

$$2\sqrt{s\cdot\alpha\cdot(1-\alpha)} \geq 4\sum_{S\subseteq[n]} |S| \left(\widehat{f}(S)\right)^2 = \mathbf{AS}(f)$$

as required. □

References

[ACR+07] Andris Ambainis, Andrew M. Childs, Ben Reichardt, Robert Spalek, and Shengyu Zhang. Any and-or formula of size n can be evaluated in time $n^{1/2+o(1)}$ on a quantum computer. In *FOCS*, pages 363–372. IEEE Computer Society, 2007.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[BJT99] Nader H. Bshouty, Jeffrey C. Jackson, and Christino Tamon. Uniform-distribution attribute noise learnability. In *Workshop on Computational Learning Theory*, pages 75–80, 1999.

[BKS98] Itai Benjamini, Gil Kalai, and Oded Schramm. Noise sensitivity of boolean functions and applications to percolation, 1998.

[FGG08] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum algorithm for the hamiltonian nand tree. *Theory of Computing*, 4(1):169–190, 2008.

[Hås98] Johan Håstad. The shrinkage exponent of de morgan formulas is 2. *SIAM J. Comput.*, 27(1):48–64, 1998.

[Khr71] V.M. Khrapchenko. A method of determining lower bounds for the complexity of π schemes. *Matematicheskie Zametki*, 10:83–92, 1971. In Russian.

[KKL88] Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions (extended abstract). In *FOCS*, pages 68–80. IEEE Computer Society, 1988.

[Kou93] E. Koutsoupias. Improvements on Khrapchenko's theorem. *Theoretical Computer Science*, 116(2):399–403, August 1993.

[Lee09] Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.

[LLS06] Sophie Laplante, Troy Lee, and Mario Szegedy. The quantum adversary method and classical formula size lower bounds. *Computational Complexity*, 15(2):163–196, 2006.

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, 1993.

[O'D02] Ryan O'Donnell. Hardness amplification within np. In John H. Reif, editor, *STOC*, pages 751–760. ACM, 2002.

[OS03] Ryan O'Donnell and Rocco A. Servedio. New degree bounds for polynomial threshold functions. In Lawrence L. Larmore and Michel X. Goemans, editors, *STOC*, pages 325–334. ACM, 2003.

[Rei09] Ben Reichardt. Span programs and quantum query complexity: The general adversary bound is nearly tight for every boolean function. In *FOCS*, pages 544–551. IEEE Computer Society, 2009.

[RS08] Ben Reichardt and Robert Spalek. Span-program-based quantum algorithm for evaluating formulas. In Cynthia Dwork, editor, *STOC*, pages 103–112. ACM, 2008.

[Weg87] Ingo Wegener. *The complexity of Boolean functions*. Wiley-Teubner, 1987.

[Zwi91] Uri Zwick. An extension of khrapchenko's theorem. *Inf. Process. Lett.*, 37(4):215–217, 1991.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de